

Glenn M, Grigor MR, Molenaar AJ, Davis SR;

XX MPI; 2003-275306/27.  
 DR N-PSDB; ABX11388.  
 XX  
 PT New histatin polynucleotides and polypeptides expressed in bovine  
 PT tissues, useful for treating bacterial or fungal infections in mammals,  
 PT e.g. infections of the oral cavity, vagina, urethra, ear or skin, or  
 PT systemic infections -  
 XX  
 PS Claim 6; Page 13; 17pp; English.  
 XX  
 CC The invention discloses isolated histatin polynucleotides and  
 CC polypeptides, which comprise sequences expressed in bovine. Histatins are  
 CC histidine rich and specific to the salivary secretions. They are believed  
 CC to function as part of the non-immune defence system, particularly in the  
 CC oral cavity and have promise as therapeutic agents in humans with oral  
 CC candidosis. Also disclosed are compositions comprising the histatin  
 CC polypeptide or polynucleotide, and at least one component consisting of  
 CC physiological or pharmaceutical carriers or immunostimulants. The  
 CC histatin polynucleotide, polypeptide, or the cosmetic composition  
 CC comprising the polypeptide, is useful for treating (e.g. gene therapy) a  
 CC disorder in a mammal, particularly microbial or fungal infections. The  
 CC histatin polynucleotide is also useful in genome mapping, physical  
 CC mapping or in the positional cloning of genes. Specifically, the  
 CC polypeptide or polynucleotide is useful for treating fungal or bacterial  
 CC infections of the oral cavity (e.g. dental caries, plaque or tartar),  
 CC vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis  
 CC caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating  
 CC systemic infections such as systemic Candida infection. The polypeptide  
 CC or polynucleotide is also useful for veterinary applications e.g. for  
 CC treating mastitis. The sequence presented is a bovine histatin like  
 CC polypeptide.  
 CC  
 XX Sequence 58 AA;  
 SO

Query Match 100.0%; Score 316; DB 24; Length 58;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-32;  
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTFFVFMALIAMIRADSSSEKRRKRRKHGRGYFOOYQYQYRPLNYPAYPP 58  
 DB 1 MKTFFVFMALIAMIRADSSSEKRRKRRKHGRGYFOOYQYQYRPLNYPAYPP 58

RESULT 2  
 ABG75794  
 ID ABG75794 standard; Protein; 58 AA.  
 XX  
 AC ABG75794;  
 XX  
 DT 01-MAY-2003 (first entry)  
 XX  
 DE Bovine histatin like polypeptide, #4.  
 XX  
 KM Cow; histatin; bovine; salivary secretion; oral cavity;  
 KM non-immune defence system; oral candidosis; gene therapy;  
 KM microbial infection; fungal infection; dental caries; plaque; tartar;  
 KM cystic fibrosis; systemic infection; Candida infection; mastitis;  
 KM fungicide; antibacterial.  
 OS Bos taurus.  
 XX  
 PN US2002164625-A1.  
 XX  
 PD 07-NOV-2002.  
 XX  
 PF 19-FEB-2002; 2002US-0079754.  
 XX  
 PR 23-AUG-1999; 99US-150330P.  
 PR 22-AUG-2000; 99US-162701P.  
 PR 27-OCT-2000; 2000US-0644190.  
 XX

PA (GENE-) GENESIS RES & DEV CORP LTD.  
 XX  
 PI Glenn M, Grigor MR, Molenaar AJ, Davis SR;  
 XX  
 DR MPI; 2003-275306/27.  
 DR N-PSDB; ABX11391.  
 XX  
 PT New histatin polynucleotides and polypeptides expressed in bovine  
 PT tissues, useful for treating bacterial or fungal infections in mammals,  
 PT e.g. infections of the oral cavity, vagina, urethra, ear or skin, or  
 PT systemic infections -  
 XX  
 PS Claim 6; Page 14; 17pp; English.  
 XX  
 CC The invention discloses isolated histatin polynucleotides and  
 CC polypeptides, which comprise sequences expressed in bovine. Histatin are  
 CC histidine rich and specific to the salivary secretions. They are believed  
 CC to function as part of the non-immune defence system, particularly in the  
 CC oral cavity and have promise as therapeutic agents in humans with oral  
 CC candidosis. Also disclosed are compositions comprising the histatin  
 CC polypeptide or polynucleotide, and at least one component consisting of  
 CC physiological or pharmaceutical carriers or immunostimulants. The  
 CC histatin polynucleotide, polypeptide, or the cosmetic composition  
 CC comprising the polypeptide, is useful for treating (e.g. gene therapy) a  
 CC disorder in a mammal, particularly microbial or fungal infections. The  
 CC histatin polynucleotide is also useful in genome mapping, physical  
 CC mapping or in the positional cloning of genes. Specifically, the  
 CC polypeptide or polynucleotide is useful for treating fungal or bacterial  
 CC infections of the oral cavity (e.g. dental caries, plaque or tartar),  
 CC vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis  
 CC caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating  
 CC systemic infections such as systemic Candida infection. The polypeptide  
 CC or polynucleotide is also useful for veterinary applications e.g. for  
 CC treating mastitis. The sequence presented is a bovine histatin like  
 CC polypeptide.  
 CC  
 XX Sequence 58 AA;  
 SO

Query Match 100.0%; Score 316; DB 24; Length 58;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-32;  
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTFFVFMALIAMIRADSSSEKRRKRRKHGRGYFOOYQYQYRPLNYPAYPP 58  
 DB 1 MKTFFVFMALIAMIRADSSSEKRRKRRKHGRGYFOOYQYQYRPLNYPAYPP 58

RESULT 3  
 ABG75792  
 ID ABG75792 standard; Protein; 58 AA.  
 XX  
 AC ABG75792;  
 XX  
 DT 01-MAY-2003 (first entry)  
 XX  
 DE Bovine histatin like polypeptide, #2.  
 XX  
 KM Cow; histatin; bovine; salivary secretion; oral cavity;  
 KM non-immune defence system; oral candidosis; gene therapy;  
 KM microbial infection; fungal infection; dental caries; plaque; tartar;  
 KM cystic fibrosis; systemic infection; Candida infection; mastitis;  
 KM fungicide; antibacterial.  
 OS Bos taurus.  
 XX  
 PN US2002164625-A1.  
 XX  
 PD 07-NOV-2002.  
 XX  
 PF 19-FEB-2002; 2002US-0079754.  
 XX  
 PR 23-AUG-1999; 99US-150330P.  
 PR 22-AUG-2000; 99US-162701P.  
 PR 29-OCT-1999; 99US-162701P.  
 XX

```

XX          22-AUG-2000; 2000US-0644190.
PR          27-OCT-2000; 2000US-0699146.
XX          (GENE-) GENESTIS RES & DEV CORP LTD.
PA          Glenn M., Grigor MR, Molenaar AJ, Davis SR,
PI          WPI: 2003-275306/27.
DR          N-PESDB; ABX11389.
XX          New histatin polynucleotides and polypeptides expressed in bovine
PT          tissues, useful for treating bacterial or fungal infections in mammals,
PT          e.g. infections of the oral cavity, vagina, urethra, ear or skin, or
XX          systemic infections -
XX          Claim 6, Page 13; 17pp; English.
XX
XX          The invention discloses isolated histatin polynucleotides and
XX          polypeptides, which comprise sequences expressed in bovine. Histatins are
XX          histidine rich and specific to the salivary secretions. They are believed
XX          to function as part of the non-immune defence system, particularly in the
XX          oral cavity and have promise as therapeutic agents in humans with oral
XX          candidosis. Also disclosed are compositions comprising the histatin
XX          polypeptide or polynucleotide, and at least one component consisting of
XX          physiological or pharmaceutical carriers or immunostimulants. The
XX          histatin polynucleotide, polypeptide, or the cosmetic composition
XX          comprising the polypeptide, is useful for treating (e.g. gene therapy) a
XX          disorder in a mammal, particularly microbial or fungal infections. The
XX          histatin polynucleotide is also useful in genome mapping, physical
XX          mapping or in the positional cloning of genes. Specifically, the
XX          polypeptide or polynucleotide is useful for treating fungal or bacterial
XX          infections of the oral cavity (e.g. dental caries, plaque or tartar),
XX          vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis
XX          caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating
XX          systemic infections such as systemic Candida infection. The polypeptide
XX          or polynucleotide is also useful for veterinary applications e.g. for
XX          treating mastitis. The sequence presented is a bovine histatin like
XX          polypeptide.
XX          Sequence      58 AA;
XX          SQ
XX
XX          Query March           99.7%; Score 315; DB 24; Length 58;
XX          Best Local Similarity 98.3%; Pred. No. 3,5e-32;
XX          Matches 57; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
QY          1 MKIFIVFVIMLILMIIRADSSSEKRRKKRKHGRGYCOYOYORYPANVPAYPPF 58
Db          1 MKIFIFIFIMLILMIIRADSSSEKRRKKRKHGRHYGOYOYORYPANVPAYPPF 58
XX
RESULT 4
ABG75796
ID ID ABG75796 standard; Protein; 59 AA.
XX
AC ABG75796;
XX
DT 01-MAY-2003 (first entry)
XX
DE Bovine histatin like polypeptide, #6.
XX
KW Cow; histatin; bovine; salivary secretion; oral cavity;
KM non-immune defence system; oral candidosis; gene therapy;
KW microbial infection; fungal infection; dental caries; plaque; tartar;
KM cystic fibrosis; systemic infection; Candida infection; mastitis;
KW fungicide; antibacterial..
XX
OS Bos taurus.
XX
PN US2002164625-A1.
XX
PD 07-NOV-2002.
XX
PF 19-FEB-2002; 2002US-0079754.
```

XX	23-AUG-1999;	99US-150330P.
PR	29-OCT-1999;	99US-162701P.
PR	22-AUG-2000;	2000US-0644190.
PR	27-OCT-2000;	2000US-0699146.
XX		
PA	(GENE-) GENESIS RES & DEV CORP LTD.	
PI	Glenn M, Grigor MR, Molenaar AJ, Davis SR;	
XX		
DR	WPI; 2003-275306/27.	
XX	N-PSDB; ABX11393.	
PT	New histatin polynucleotides and polypeptides expressed in bovine	
PT	tissues, useful for treating bacterial or fungal infections in mammals,	
PT	e.g. infections of the oral cavity, vagina, urethra, ear or skin, or	
XX	systemic infections	
PS		
XX	Claim 6; Page 14; 17pp; English.	
XX		
CC	The invention discloses isolated histatin polynucleotides and	
CC	polypeptides, which comprise sequences expressed in bovine. Histatins are	
CC	histidine rich and specific to the salivary secretions. They are believed	
CC	to function as part of the non-immune defence system, particularly in the	
CC	oral cavity and have promise as therapeutic agents in humans with oral	
CC	candidosis. Also disclosed are compositions comprising the histatin	
CC	polypeptide or polynucleotide, and at least one component consisting of	
CC	physiological or pharmaceutical carriers or immunostimulants. The	
CC	histatin polynucleotide, polypeptide, or the cosmetic composition	
CC	comprising the polypeptide, is useful for treating (e.g. gene therapy) a	
CC	disorder in a mammal, particularly microbial or fungal infections. The	
CC	histatin polynucleotide is also useful in genome mapping, physical	
CC	mapping or in the positional cloning of genes. Specifically, the	
CC	polypeptide or polynucleotide is useful for treating fungal or bacterial	
CC	infections of the oral cavity (e.g. dental caries, plaque or tartar),	
CC	vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis	
CC	caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating	
CC	systemic infections such as systemic Candida infection. The polypeptide	
CC	or polynucleotide is also useful for veterinary applications e.g. for	
CC	treating mastitis. The sequence presented is a bovine histatin like	
CC	polypeptide.	
XX		
SO	Sequence 59 AA;	
Query Match	53.8%; Score 170; DB 24; Length 59;	
Best Local Similarity	100.0%; Pred. No. 5.8e-14;	
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 MKRIFPFIALLIAMRADSSSEKRRKRRKKKH 34	
DB	1 MKRIFPFIALLIAMRADSSSEKRRKRRKKKH 34	
RESULT 5		
ID	ABG75793 standard; Protein; 70 AA.	
XX	ABG75793	
XX	ABG75793;	
DT	01-MAY-2003 (first entry)	
DE	Bovine histatin like polypeptide, #3.	
XX		
KW	Cow; histatin; bovine; salivary secretion; oral cavity;	
KW	non-immune defence system; oral candidosis; gene therapy;	
KW	microbial infection; fungal infection; dental caries; plaque; tartar;	
KW	cystic fibrosis; systemic infection; Candida infection; mastitis;	
KW	fungicide; antibacterial.	
OS	Bos taurus.	
XX		
XX	US2002164625-A1.	
XX		

PD 07-NOV-2002.  
 XX 19-FEB-2002; 2002US-0079754.  
 PF 23-AUG-1999; 99US-150330P.  
 PR 23-OCT-1999; 99US-162701P.  
 PR 22-AUG-2000; 2000US-0644190.  
 PR 27-OCT-2000; 2000US-0699146.  
 XX (GENE-) GENESIS RES & DEV CORP LTD.  
 PA Glenn M, Grigor MR, Molenaar AJ, Davis SR;  
 XX WPI; 2003-275306/27.  
 DR N-PADB; ABX11390.  
 XX  
 PT New histatin polynucleotides and polypeptides expressed in bovine  
 PT tissues, useful for treating bacterial or fungal infections in mammals,  
 PT e.g. infections of the oral cavity, vagina, urethra, ear or skin, or  
 PT systemic infections -  
 XX  
 PS Claim 6; Page 13-14; 17pp; English.  
 XX  
 CC The invention discloses isolated histatin polynucleotides and  
 CC polypeptides, which comprise sequences expressed in bovine. Histatins are  
 CC histidine rich and specific to the salivary secretions. They are believed  
 CC to function as part of the non-immune defence system, particularly in the  
 CC oral cavity and have promise as therapeutic agents in humans with oral  
 CC candidosis. Also disclosed are compositions comprising the histatin  
 CC polypeptide or polynucleotide, and at least one component consisting of  
 CC physiological or pharmaceutical carriers or immunostimulants. The  
 CC histatin polynucleotide, polypeptide, or the cosmetic composition  
 CC comprising the polypeptide, is useful for treating (e.g. gene therapy) a  
 CC disorder in a mammal, particularly microbial or fungal infections. The  
 CC histatin polynucleotide, is also useful in genome mapping, physical  
 CC mapping or in the positional cloning of genes. Specifically, the  
 CC polypeptide or polynucleotide is useful for treating fungal or bacterial  
 CC infections of the oral cavity (e.g. dental caries, plaque or tartar),  
 CC vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis  
 CC caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating  
 CC systemic infections such as systemic Candida infection. The polypeptide  
 CC or polynucleotide is also useful for veterinary applications e.g. for  
 CC treating mastitis. The sequence presented is a bovine histatin like  
 CC polypeptide.  
 CC  
 XX  
 SQ Sequence 70 AA;  
 Query Match 53.5%; Score 169; DB 24; Length 70;  
 Best Local Similarity 97.1%; Pred. No. 9.4e-14;  
 Matches 33; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKTFFIFIMALLIAMIADSEKRRKRRKHH 34  
 Db 1 MKTFFIFIMALLIAMIADSEKRRKRRKHH 34  
 ID ABG75795 standard; Protein; 21 AA.  
 AC ABG75795;  
 XX  
 DT 01-MAY-2003 (first entry)  
 XX  
 DE Bovine histatin like polypeptide, #5.  
 XX  
 KW Cow; histatin; bovine; salivary secretion; oral cavity;  
 KW non-immune defence system; oral candidosis; gene therapy;  
 KW microbial infection; fungal infection; dental caries; plaque; tartar;  
 KW cystic fibrosis; systemic infection; Candida infection; mastitis;  
 KW fungicide; antibacterial.  
 XX  
 OS Bos taurus.

XX US2002164625-A1.  
 PN 07-NOV-2002.  
 XX  
 PD 19-FEB-2002; 2002US-0079754.  
 PF 23-AUG-1999; 99US-150330P.  
 PR 23-OCT-1999; 99US-162701P.  
 PR 22-AUG-2000; 2000US-0644190.  
 PR 27-OCT-2000; 2000US-0699146.  
 XX (GENE-) GENESIS RES & DEV CORP LTD.  
 PA Glenn M, Grigor MR, Molenaar AJ, Davis SR;  
 XX WPI; 2003-275306/27.  
 DR N-PADB; ABX11392.  
 XX  
 PT New histatin polynucleotides and polypeptides expressed in bovine  
 PT tissues, useful for treating bacterial or fungal infections in mammals,  
 PT e.g. infections of the oral cavity, vagina, urethra, ear or skin, or  
 PT systemic infections -  
 XX  
 PS Claim 6; Page 14; 17pp; English.  
 XX  
 CC The invention discloses isolated histatin polynucleotides and  
 CC polypeptides, which comprise sequences expressed in bovine. Histatins are  
 CC histidine rich and specific to the salivary secretions. They are believed  
 CC to function as part of the non-immune defence system, particularly in the  
 CC oral cavity and have promise as therapeutic agents in humans with oral  
 CC candidosis. Also disclosed are compositions comprising the histatin  
 CC polypeptide or polynucleotide, and at least one component consisting of  
 CC physiological or pharmaceutical carriers or immunostimulants. The  
 CC histatin polynucleotide, polypeptide, or the cosmetic composition  
 CC comprising the polypeptide, is useful for treating (e.g. gene therapy) a  
 CC disorder in a mammal, particularly microbial or fungal infections. The  
 CC histatin polynucleotide is also useful in genome mapping, physical  
 CC mapping or in the positional cloning of genes. Specifically, the  
 CC polypeptide or polynucleotide is useful for treating fungal or bacterial  
 CC infections of the oral cavity (e.g. dental caries, plaque or tartar),  
 CC vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis  
 CC caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating  
 CC systemic infections such as systemic Candida infection. The polypeptide  
 CC or polynucleotide is also useful for veterinary applications e.g. for  
 CC treating mastitis. The sequence presented is a bovine histatin like  
 CC polypeptide.  
 CC  
 XX  
 SQ Sequence 21 AA;  
 Query Match 35.4%; Score 112; DB 24; Length 21;  
 Best Local Similarity 94.7%; Pred. No. 3.7e-07;  
 Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 38 FQGYQPYRXYPLNYPYPAYP 56  
 Db 1 FQGYQPYRXYPLNYPYPAYP 19  
 ID AAY94527 standard; protein; 62 AA.  
 AC AAY94527;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human statherin protein.  
 XX  
 KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;  
 KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;  
 KW aetna; allergy; diabetes mellitus; fungal; bacterial infection;  
 KW cancer; leukemia; adenocarcinoma; melanoma.



```
XX OS Homo sapiens.
XX PN WO200024779-A1.
XX PD 04-MAY-2000.
XX PF 22-OCT-1999; 99WO-US24046.
XX PR 23-OCT-1998; 98US-0155209.
XX PA (INCY-) INCYTE PHARM INC.
XX PI Tang YT, Corley NC, Guegler KJ, Patterson C;
XX DR WPI; 2000-350699/30.
XX DR N-PSDB; AAA48964.
XX PT Purified polypeptide used for treating or preventing a disorder
XX PT characterized by expression or activity of lysine-rich statherin
XX PT proteins -
XX PS Disclosure; Page 70; 75pp; English.
XX CC The present invention relates to human lysine-rich statherin protein
XX CC (LRSP) (AA94526). The cDNA sequence encoding this protein was identified
XX CC through analysis of a cDNA library of breast tumour tissue (BRSTN0114).
XX CC The LRSP sequence was found to have homology with human statherin (the
XX CC present sequence) and human basic histidine-rich protein (AA94528).
XX CC Human statherin is a phosphoprotein that acts as an inhibitor of
XX CC precipitation of calcium phosphate salts in the oral cavity. The LRSP
XX CC polypeptide and its antagonists may be useful for treating or preventing
XX CC disorders associated with the activity of LRSP. Such disorders include
XX CC autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
XX CC diabetes mellitus), bacterial and fungal infection and cancers (such as
XX CC leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be useful
XX CC for diagnosis of the above disorders.
XX SQ Sequence 62 AA;
XX
XX Query Match 34.8%; Score 110; DB 21; Length 62;
XX Best Local Similarity 49.1%; Pred. No. 2.1e-06;
XX Matches 27; Conservative 7; Mismatches 19; Indels 2; Gaps 2;
XX
OY 1 MKIFVFVFMALILAMIRADSESEKRRKRRKHHGYYQQVQYQRYPLNYPAY 55
DB 1 MKIFVFALILALMSMTGADSHAKRHGKRRKFKHSHRGYSNY 47
XX
XX RESULT 8
XX AA94528
XX ID AA94528 standard; protein; 51 AA.
XX AC AAY94528;
XX XX
XX DT 06-OCT-2000 (first entry)
XX DE Human basic histidine-rich protein.
XX KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
XX KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
XX KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;
XX KW cancer; leukemia; adenocarcinoma; melanoma.
XX OS Homo sapiens.
XX PN WO200024779-A1.
XX PD 04-MAY-2000.
XX PF 22-OCT-1999; 99WO-US24046.
XX PR 23-OCT-1998; 98US-0155209.
```

```
XX PA (INCY-) INCYTE PHARM INC.
XX PI Tang YT, Corley NC, Guegler KJ, Patterson C;
XX DR WPI; 2000-350699/30.
XX DR N-PSDB; AAA48965.
XX PT Purified polypeptide used for treating or preventing a disorder
XX PT characterized by expression or activity of lysine-rich statherin
XX PT proteins -
XX PS Disclosure; Page 71; 75pp; English.
XX CC The present invention relates to human lysine-rich statherin protein
XX CC (LRSP) (AA94526). The cDNA sequence encoding this protein was identified
XX CC through analysis of a cDNA library of breast tumour tissue (BRSTN0114).
XX CC The LRSP sequence was found to have homology with human statherin
XX CC (AA94527) and human basic histidine-rich protein (the present
XX CC sequence). Human statherin is a phosphoprotein that acts as an inhibitor
XX CC of precipitation of calcium phosphate salts in the oral cavity. The LRSP
XX CC polypeptide and its antagonists may be useful for treating or preventing
XX CC disorders associated with the activity of LRSP. Such disorders include
XX CC autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
XX CC diabetes mellitus), bacterial and fungal infection and cancers (such as
XX CC leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be useful
XX CC for diagnosis of the above disorders.
XX SQ Sequence 51 AA;
XX
XX Query Match 30.1%; Score 95; DB 21; Length 51;
XX Best Local Similarity 46.8%; Pred. No. 0.00013;
XX Matches 22; Conservative 6; Mismatches 13; Indels 6; Gaps 2;
XX
OY 1 MKIFVFVFMALILAMIRADSESEKRRKRRKHHGYYQQV 41
DB 1 MKIFVFALILALMSMTGADSHAKRHGKRRKFKHSHRGYSNY 47
XX
XX RESULT 9
XX AAM23999
XX ID AAM23999 standard; protein; 51 AA.
XX AC AAM23999;
XX XX
XX DT 12-OCT-2001 (first entry)
XX DE Human EST encoded protein; SEQ ID NO: 1524.
XX KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
XX KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
XX KW diagnostics; forensic test; gene mapping; genetic disorder;
XX KW biodiversity; gene therapy; nutrition.
XX OS Homo sapiens.
XX PN WO200154477-A2.
XX PD 02-AUG-2001.
XX PF 25-JAN-2001; 2001WO-US02687.
XX PR 25-JAN-2000; 2000US-0491404.
XX PR 17-JUL-2000; 2000US-0617746.
XX PR 03-AUG-2000; 2000US-0631451.
XX PR 15-SEP-2000; 2000US-0663870.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX PI Cao Y, Dermanac RA, Zhang J, Werhman T;
XX DR WPI; 2001-476164/51.
```

DR N-PSDB; AAH98658.  
XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use -  
XX  
PS Claim 20; Page 1051; 1275pp; English.  
XX The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea (ESTs)  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensic, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention.  
XX  
SQ Sequence 51 AA;  
Query Match 30.1%; Score 95; DB 22; Length 51;  
Best Local Similarity 46.8%; Pred. No. 0.00013;  
Matches 22; Conservative 6; Mismatches 13; Indels 6; Gaps 2;  
QY 1 MKFIFVFMALILAMIRADSSSEKRLRRKKK-----HRCYFOQY 41  
DB 1 MKFVFALILALMLMTGADSHAKRHGKRRKFKHSHRGYSNY 47  
RESULT 10  
AAU90983  
ID AAU90983 standard; Peptide; 51 AA.  
AC AAU90983;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Transplant media associated antimicrobial peptide #19.  
XX  
KM Transplant; antimicrobial peptide; pore forming agent;  
KM cell surface receptor binding compound; kidney transplant;  
KM cardioplegia; organ transplant; transplant rejection.  
XX  
OS Homo sapiens.  
XX  
PN WO200209738-A1.  
XX  
PD 07-FEB-2002.  
XX  
PF 27-JUL-2001; 2001WO-US23785.  
XX  
PR 28-JUL-2000; 2000US-221632P.  
PR 17-NOV-2000; 2000US-249602P.  
PR 15-MAY-2001; 2001US-290932P.  
XX  
PA (MURP/) MURPHY C J.  
XX  
PI Murphy CJ, Reid TW, Mcanulty JF;  
XX  
DR WPI; 2002-268995/31.  
XX  
PT Media comprising antimicrobial polypeptides or pore forming agents  
PT and/or cell surface receptor binding compounds useful for the storage  
PT and preservation of organs prior to transplant -  
XX  
PS Disclosure; Page 26; 78pp; English.  
XX The invention describes new transplant compositions comprising  
CC antimicrobial polypeptides or pore forming agents and/or cell surface  
CC receptor binding compounds. The media is capable of extending the  
CC preservation period past 72 hours and can provide organs with increased  
CC functionality upon transplant. animals receiving kidneys scored in the  
CC media of the present invention for either three or four days had serum  
CC creatinine levels of less than half of those observed in control animals  
CC receiving kidneys stored in UW solution (defined in the specification)

CC alone. Lower serum creatinine levels are indicative of healthier kidneys  
CC and a more preferable prognosis for the transplant patient. The media of  
CC the invention are useful for decreasing the incidence and/or severity of  
CC delayed graft function in patients receiving transplanted kidneys scored  
CC and/or treated in the media. The media may also be used in procedures  
CC such as cardioplegia. It is contemplated that transplant of healthier  
CC organs leads to a decrease in chronic rejection. This sequence represents  
CC an antimicrobial peptide studied in the development of the transplant  
CC media.  
XX  
SQ Sequence 51 AA;  
Query Match 30.1%; Score 95; DB 23; Length 51;  
Best Local Similarity 46.8%; Pred. No. 0.00013;  
Matches 22; Conservative 6; Mismatches 13; Indels 6; Gaps 2;  
QY 1 MKFIFVFMALILAMIRADSSSEKRLRRKKK-----HRCYFOQY 41  
DB 1 MKFVFALILALMLMTGADSHAKRHGKRRKFKHSHRGYSNY 47  
RESULT 11  
ABR48455  
ID ABR48455 standard; Protein; 78 AA.  
AC ABR48455;  
XX  
DT 13-JUN-2003 (first entry)  
XX  
DE Human Chimerin.  
XX  
KM Human; GENSET; therapeutic; therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200294864-A2.  
XX  
PD 28-NOV-2002.  
XX  
PF 06-AUG-2001; 2001WO-IB01715.  
XX  
PR 25-MAY-2001; 2001US-293574P.  
PR 15-JUN-2001; 2001US-298698P.  
PR 29-JUN-2001; 2001US-302277P.  
PR 13-JUL-2001; 2001US-305456P.  
XX  
PA (GENST ) GENSET.  
XX  
PI Benjamin S, Tanaka H;  
XX  
DR WPI; 2003-129412/12.  
XX  
DR N-PSDB; ACCS1062.  
XX  
PT New GENSET polynucleotides and polypeptides, useful for preparing a  
PT composition for treating GENSET-related disorders and as reagents in  
PT assays to quantitatively determined levels of GENSET expression in  
PT biological samples -  
XX  
PS Claim 2; Page 405; 505pp; English.  
XX  
XX The present invention relates to novel human GENSET coding sequences  
CC (ACCS1060-ACCS1115) and proteins (ABR48453-ABR48508). The GENSET  
CC sequences are useful for preparing a composition for treating  
CC GENSET-related disorders. They can also be used as markers for tissues in  
CC which the corresponding protein is preferentially expressed, as molecular  
CC weight markers on Southern gels, as chromosome markers or tags to  
CC identify chromosomes, and as reagents in assays to quantitatively  
CC determined levels of GENSET expression in biological samples.  
XX  
SQ Sequence 78 AA;  
Query Match 29.9%; Score 94.5; DB 24; Length 78;  
Best Local Similarity 32.7%; Pred. No. 0.00024;



PN W0200259260-A2.  
XX  
PD 01-AUG-2002.  
XX  
PF 16-NOV-2001; 2001WO-US42950.  
XX  
PR 17-NOV-2000; 2000US-0714936.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Goodrich RW, Liu C, Zhou P, Aundi V, Zhang J, Zhao QA,  
PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX  
DR WPI: 2002-590824/63.  
DR N-PSDB; ABQ99552.  
XX  
PT New isolated polynucleotide, useful in research, diagnostic or  
PT therapeutic methods, e.g. preventing or treating disorders involving  
PT aberrant protein expression or biological activity -  
XX  
PS Claim 20; SEQ ID 626; 394pp; English.

CC The present invention relates to novel human coding sequences  
CC (ABQ99268-ABQ99608) and proteins (ABP64682-ABP65022). The sequences are  
CC useful in therapeutic, diagnostic and research methods. The  
CC polynucleotides may be used in the field of molecular biology as  
CC hybridisation probes, primers for PCR, for chromosome and gene mapping,  
CC for the recombinant production of protein, or in generation of anti-sense  
CC DNA or RNA. The polynucleotides are useful in diagnostics as expressed  
CC sequence tags (ESTs) for identifying expressed genes or for physical  
CC mapping of the human genome. The proteins may be used as molecular weight  
CC markers, or as nutritional sources or supplements. The proteins may be  
CC used to maintain and expand cell population in a totipotent or  
CC pluripotent state useful for re-engineering damaged or diseased  
CC tissues, transplantation, manufacture of bio-pharmaceuticals or the  
CC development of bio-sensors. The polynucleotides and proteins are useful  
CC for preventing, treating or ameliorating disorders involving aberrant  
CC protein expression or biological activity, e.g. haematopoietic disorders,  
CC central/peripheral nervous system diseases, mechanical and traumatic  
CC disorders, non-healing wounds, immune deficiencies and disorders,  
CC infectious diseases caused by viral, bacterial or fungal infection,  
CC autoimmune disorders, allergic reactions and conditions, coagulation  
CC disorders, or cancer. The polynucleotide sequences of the invention were  
CC assembled from ESTs isolated mainly by sequencing by hybridisation, and  
CC in some cases, sequences obtained from one or more public databases.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 62 AA;

Query Match 27.2%; Score 86; DB 23; Length 62;  
Best Local Similarity 42.1%; Pred. No. 0.0022;  
Matches 24; Conservative 5; Mismatches 26; Indels 2; Gaps 1;

QY 1 MKRIFPIMLILAMIRADSSSEKRRKRRKHHRGV--FQYQYQRYPLNPPAY 55  
DB 1 MKRIFPIMLILAMIRADSSSEKRRKRRKHHRGV--FQYQYQRYPLNPPAY 57

RESULT 15  
ABB69931  
ID ABB69931 standard; Protein: 192 AA.  
XX  
AC ABB69931;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 36585.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.

OS Drosophila melanogaster.  
XX  
PN W0200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW,  
PI N-PSDB; ABL14034.  
XX  
DR WPI: 2001-656860/75.  
DR N-PSDB; ABL14034.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure, SEQ ID NO 36585, 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 192 AA;

Query Match 23.7%; Score 75; DB 22; Length 192;  
Best Local Similarity 27.4%; Pred. No. 0.18; Length 192;  
Matches 23; Conservative 12; Mismatches 15; Indels 34; Gaps 6;

QY 7 VFMALILAMIRAD-----SSEKRRKRRKHHRGV--FQYQYQRYPLNPPAY 40  
DB 10 LVVLLQYEVVRDSDSSSSSGEKRRKRRKHHRGV--FQYQYQRYPLNPPAY 69  
QY 41 --YQYQRYPLNPPAY--FQYQYQRYPLNPPAY 57  
DB 70 YLYNPPYMPY-PYQWGMGYPSYPY 92

Search completed: December 13, 2003, 03:07:17  
Job time : 49 secs

```
QY      1 MKLFI FVF- I MALLIAMI RADSSSEKRRKKKHGRGFOOQORPQRYPPLNY -PRAYPF 57  
       ||| : | | : : | : | : | : | : | : |  
Db      1 MKSFLVNNALATLPFLAVEONOKOPACHENDERPFYOXTAP--VPMYVVPNISRY 57
```

## RESULT 2

US-08-108-883-2  
Sequence 2, Application US/08308863  
Patent No. 5576300  
GENERAL INFORMATION:  
APPLICANT: Mukerji, P.  
APPLICANT: Prieto, P. A.  
APPLICANT: Seo, A. E.-Y.  
APPLICANT: Baxter, J. H.  
APPLICANT: Cummings, R.D.  
TITLE OF INVENTION: Method for inhibition of Human Rotavirus Infection.  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lonnie R. Drayer  
ADDRESSEE: Ross Products Division  
ADDRESSEE: Abbott Laboratories  
STREET: 625 Cleveland Avenue  
CITY: Columbus  
STATE: Ohio  
COUNTRY: United States  
ZIP: 43215  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh System 7.1  
SOFTWARE: ClarisWorks 1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/308,883  
FILING DATE: 16-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA: No. 5576300 applicable  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (614) 624-3774  
TELEFAX: (614) 624-3074  
TELEX: No. 5576300e  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 182 amino acids  
TYPE: Amino acid  
STRANDEDNESS:  
TOPOLOGY: Linear  
MOLECULE TYPE: Protein.  
DESCRIPTION:  
HYPOTHETICAL:  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM:  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:  
ORGANELLE:  
IMMEDIATE SOURCE:  
LIBRARY:  
CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: L. Hansson et al  
TITLE: DNA Encoding Kappa-Casein,  
JOURNAL: Process for Obtaining the Protein and Use The

## VOLUME:

ISSUE:

PAGES:

DATE:

DOCUMENT NUMBER: PCT/WO93/15196

FILING DATE: 25-JAN-1993

PUBLICATION DATE: 05-AUG-1993

RELEVANT RESIDUES IN SEQ ID NO:

US-08-108-883-2

Query Match

19.9%; Score 63; DB 1; Length 182;

Best Local Similarity 28.8%; Pred. No. 1.9;

Matches 17; Conservative 13; Mismatches 25; Indels 4; Gaps 3;

Qy 1 MKIFIFVF-IMALILMIRADSESEKRRKRGHGGYQFOYQYORPLNY-PYAPF 57  
Db 1 MRSFLVYNAALTLTFLPLAVEVONQOPACHENDERPFYOKTAPY--VPMYVFNSTPY 57

## RESULT 3

US-08-730-163-2  
Sequence 2, Application US/08730163  
Patent No. 5712250  
GENERAL INFORMATION:  
APPLICANT: Mukerji, P.  
APPLICANT: Prieto, P. A.  
APPLICANT: Seo, A. E.-Y.  
APPLICANT: Baxter, J. H.  
APPLICANT: Cummings, R.D.  
TITLE OF INVENTION: Product for inhibition of Human Rotavirus Infection.  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lonnie R. Drayer  
ADDRESSEE: Ross Products Division  
ADDRESSEE: Abbott Laboratories  
STREET: 625 Cleveland Avenue  
CITY: Columbus  
STATE: Ohio  
COUNTRY: United States  
ZIP: 43215  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb storage (B)COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh System 7.1(D)SOFTWARE: ClarisWorks 1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/730,163  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/308,882  
FILING DATE: 16-SEP-1994  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (614) 624-3774  
TELEFAX: (614) 624-3074  
TELEX: No. 5712250e  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 182 amino acids  
TYPE: Amino acid  
STRANDEDNESS:  
TOPOLOGY: Linear  
MOLECULE TYPE: Protein.  
DESCRIPTION:  
HYPOTHETICAL:  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM:  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:

```

1 CELL LINE:
2 ORGANELLE:
3 IMMEDIATE SOURCE:
4 LIBRARY:
5 CLONE:
6 POSITION IN GENOME:
7 CHROMOSOME/SEGMENT:
8 MAP POSITION:
9 UNITS:
10 FEATURE:
11 NAME/KEY:
12 LOCATION:
13 IDENTIFICATION METHOD:
14 OTHER INFORMATION:
15 PUBLICATION INFORMATION:
16 AUTHORS: L. Hansson et al
17 TITLE: DNA Encoding Kappa-Casein, Process for Obtaining the Protein and Use The
18 JOURNAL:
19 VOLUME:
20 ISSUE:
21 PAGES:
22 DATE:
23 DOCUMENT NUMBER: PCT/MO93/15196
24 FILING DATE: 25-JAN-1993
25 PUBLICATION DATE: 05-AUG-1993
26 RELEVANT RESIDUES IN SEQ ID NO:
27 JS-08-730-163-2

```

Query Match	19.9%;	Score 63;	DB 1;	Length 182;
Best Local Similarity	28.8%;	Pred. No. 1.9;		
Matches	17;	Conservative	13;	Mismatches 25;
			Indels	4;
			Gaps	3

```
QY      1 MKIIFVFF-IMALLAMIRADSSSEKKRRKKKHHRGVFQQQVPORYPLNY-PPAYPF 57
       |||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db      1 MKSFLVNALATLPLFLAVEVQNKQPACHENDERPEYQTAPY--VEMYYVENSYPY 57
```

RESULT 4  
US-08-256-799-2

Patent No. 6222094

GENERAL INFORMATION:

APPLICANT: HANSSON, Lennart

APPLICANT: STROEMQVIST, Mats

APPLICANT: BERGSTROM, Sven

APPLICANT: HERNELL, Olle

APPLICANT: Toernell, Jan

TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS FOR

TITLE OF INVENTION: OBTAINING THE PROTEIN AND USE THEREOF

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/256,799

FILING DATE: 06-DEC-1994

CLASSIFICATION: A35

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DK 88/92

FILING DATE: 23-JAN-1992

ATTORNEY/AGENT INFORMATION:

NAME: COOPER, Iver P.

REGISTRATION NUMBER: 28,005

REFERENCE/DOCKET NUMBER: HANSSON=1

```

?
? TELECOMMUNICATION INFORMATION
?
? TELEPHONE: 202-628-5197
?
? TELEFAX: 202-737-5528
?
? INFORMATION FOR SEQ ID NO: 2:
?
? SEQUENCE CHARACTERISTICS
?
? LENGTH: 182 amino acids
?
? TYPE: amino acid
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: protein
?
US-08-256-799-2

```

Query Match	19.9%	Score 63;	DB 3;	Length 182;
Best Local Similarity	28.8%;	Pred. No. 1.9;		
Matches 17;	Conservative 13;	Mismatches 25;	Indels 4;	Gaps 3

QY 1 MKIFIFVF-IMALILAMIRADSSSEKKRKKKHHNGYFQOQYQYRRLPLNY--PRAYPF 57

Db 1 MKSFLVNNALTLPRFLAVEVQNGQKQACHENDERPRFYQYKAPY--VPMYYVNSYPR 57

RESULT 5  
US-08-462-437-2

; Sequence 2, Application US/08462437  
 ; Patent No. 6232094  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HANSSON, Lemnatt  
 ; APPLICANT: STROMOVIST, Mats  
 ; APPLICANT: BERGSTROM, Sven  
 ; APPLICANT: HERNELL, Olle  
 ; APPLICANT: TOERNELL, Jan  
 ; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN  
 ; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF  
 ; NUMBER OF SEQUENCES: 31  
 ; CORRESPONDENCE ADDRESS:

```

; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
;

```

```

1      COMPUTER READABLE FORM:
2      MEDIUM TYPE: Floppy disk
3      COMPUTER: IBM PC compatible
4      OPERATING SYSTEM: PC-DOS/MS-DOS
5      SOFTWARE: Patent In Release #1.0, Version #1.25
6      CURRENT APPLICATION DATA:
7      APPLICATION NUMBER: US/08/462,437
8      FILING DATE: 05-JUN-1995
9      PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: DK 88/92  
FILING DATE: 23-JAN-1992  
ATTORNEY/AGENT INFORMATION:

```

;      REGISTRATION NUMBER: 28,005
;      REFERENCE/DOCKET NUMBER: HANSSON=1A
;      TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633

```

: INFORMATION FOR SEQ ID NO: 2:
:
: SEQUENCE CHARACTERISTICS:
:
:   LENGTH: 182 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: OS-08-462-437-2

```

Query Match	19.9%	Score 63;	DB 3;	Length 182;
Similarity	28.8%;	Pred. NO. 1.9;		
Best Local Match	17;	Conservative 13;	Mismatches 25;	Indels 4; Gaps 3
0y	1	MKIFLFW-IMALLTAMIRADSSSEKKRRKKKHGRGVFOOQYORPYRLNYY-PPAYPF	57	
	:	:   :	:	:   ::       :       :





```
RESULT 10
US-09-252-991A-27348
; Sequence 27348, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27348
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27348

Query Match      18.7% Score 59; DB 4; Length 225;
Best Local Similarity 33.3%; Pred. No. 7.5;
Matches 18; Conservative 7; Mismatches 9; Indels 20; Gaps 3;

Qy 18 RADSEERKRRKRGH-----RGYFOQYQYQRYPLNPPA 54
Db 124 RAD--RQRHRRRRYHFRADRGCRGPAODSPGQCPQEQAGQAEPAEDPPA 174

RESULT 11
US-09-252-991A-21881
; Sequence 21881, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21881
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21881

Query Match      18.7% Score 59; DB 4; Length 720;
Best Local Similarity 29.8%; Pred. No. 27;
Matches 14; Conservative 8; Mismatches 15; Indels 10; Gaps 1;

Qy 18 RADSEERKRRKRGH-----RGYFOQYQYQRYPLNPPA 54
Db 325 RAGQGPQRHRRQRHHRGCRGPAQGRPAGTGQRGQPARQPERPPA 371

RESULT 12
US-09-252-991A-27419
; Sequence 27419, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
```

```
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27419
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27419

Query Match      18.5% Score 58.5; DB 4; Length 247;
Best Local Similarity 39.4%; Pred. No. 9.5;
Matches 13; Conservative 6; Mismatches 11; Indels 3; Gaps 1;

Qy 18 RADSEERKRRKRGHGRGYFOQYQYQRYPLNPPA 50
Db 60 RRDTCGRHRRHRRRLDHRHRRQQQP---HPLN 89

RESULT 13
US-09-252-991A-20994
; Sequence 20994, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20994
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20994

Query Match      18.5% Score 58.5; DB 4; Length 333;
Best Local Similarity 38.7%; Pred. No. 13;
Matches 12; Conservative 5; Mismatches 11; Indels 3; Gaps 1;

Qy 25 KRHRKRK---KHRRGYFOQYQYQRYPLNPPA 52
Db 114 RRHHRPPQGTTHRRRLRQGLPVRRLPADHP 144

RESULT 14
US-09-252-991A-30311
; Sequence 30311, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30311
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
```

US-09-252-991A-30311

Query Match 18.4%; Score 58; DB 4; Length 521;  
Best Local Similarity 36.7%; Pred. No. 25;  
Matches 18; Conservative 6; Mismatches 21; Indels 4; Gaps 2;

Oy 12 LILAMIRADSE--EKRRKKKRGYFOQYQRYPLNYP--PAYPP 58  
45 LIRRTIAAGVEIVHRHRRLHQLRQLRQLRQ--RHRPAHQRP 91

RESULT 15

US-09-196-270-4  
; Sequence 4, Application US/09196270  
; Patent No. 6500636

GENERAL INFORMATION:

APPLICANT: Hecht, Peter  
APPLICANT: Madden, Kevin

TITLE OF INVENTION: CHIMERIC PRE-ACTIVATED TRANSCRIPTION

FILE REFERENCE: 50078/004002

CURRENT APPLICATION NUMBER: US/09/196,270

EARLIER FILING DATE: 1998-11-19

EARLIER FILING DATE: 1997-11-19

EARLIER FILING DATE: 1997-11-21

EARLIER FILING DATE: 1997-11-24

NUMBER OF SEQ ID NOS: 24

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4

LENGTH: 585

TYPE: PRT

ORGANISM: Yarrowia lipolytica

US-09-196-270-4

Query Match 18.4%; Score 58; DB 4; Length 585;  
Best Local Similarity 34.9%; Pred. No. 28;  
Matches 15; Conservative 6; Mismatches 18; Indels 4; Gaps 2;

Oy 19 ADSSEKRRKKRGYFOQYQRYPLNYP--PAYPP 58  
235 ADNEGAHNAVAKPMQHTHQQQQQRY--MQYPTVAGVEYP 276

Search completed: December 13, 2003, 03:09:38  
Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using bw model

Run on: December 13, 2003, 03:07:25 ; Search time 276 Seconds  
(without alignments)  
39.084 Million cell updates/sec

Title: US-10-079-754A-10

Perfect score: 316  
Sequence: 1 MKRIFVFVIMLILMIRAD.....QQQYQYRYPLNYPAYPP 58

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	316	100.0	58	14	US-10-079-754A-7
2	316	100.0	58	14	US-10-079-754A-10
3	316	99.7	58	14	US-10-079-754A-8
4	170	53.8	59	14	US-10-079-754A-12
5	169	53.5	70	14	US-10-079-754A-9
6	112	35.4	21	14	US-10-079-754A-11
7	95	29.1	51	9	US-09-917-340-19
8	94.5	29.9	78	11	US-09-992-600A-6
9	94.5	29.9	78	11	US-09-924-340-6
10	94.5	29.9	78	12	US-09-993-095B-6
11	94.5	29.9	78	12	US-10-154-678-6
12	94.5	29.9	78	12	US-09-993-570-6
13	94.5	29.9	78	15	US-10-000-489-6
14	94.5	29.9	78	15	US-10-000-986-6
15	63	19.9	103	15	US-10-142-077-33

16	63	19.9	182	12	US-10-190-115-50	Sequence 50, Appl
17	63	19.9	182	12	US-10-190-115-51	Sequence 51, Appl
18	63	19.9	182	12	US-10-190-115-52	Sequence 52, Appl
19	60.5	19.1	504	8	US-08-980-068B-21	Sequence 21, Appl
20	58	18.4	205	12	US-09-949-029-76	Sequence 76, Appl
21	58	18.4	539	15	US-10-106-698-6386	Sequence 6386, Ap
22	58	18.4	585	9	US-09-922-217-1067	Sequence 1067, Ap
23	58	18.4	585	10	US-09-833-263-1067	Sequence 1067, Ap
24	58	18.4	585	14	US-10-025-380-1067	Sequence 1067, Ap
25	57.5	18.2	1111	9	US-09-815-242-12955	Sequence 12955, A
26	56.5	17.9	476	12	US-10-153-244-100	Sequence 100, App
27	56.5	17.9	476	12	US-10-210-152-17	Sequence 17, Appl
28	56.5	17.9	535	12	US-10-225-068-26	Sequence 26, Appl
29	56.5	17.9	638	12	US-10-303-664A-36	Sequence 36, Appl
30	56.5	17.9	638	15	US-10-025-950-2	Sequence 2, Appl
31	56.5	17.9	1619	12	US-10-362-978-4	Sequence 4, Appl
32	56	17.7	118	12	US-10-024-298A-31	Sequence 31, Appl
33	56	17.7	118	12	US-10-042-211A-31	Sequence 31, Appl
34	56	17.7	137	12	US-10-024-298A-29	Sequence 29, Appl
35	56	17.7	137	12	US-09-903-190-109	Sequence 109, App
36	56	17.7	137	12	US-10-042-211A-29	Sequence 29, Appl
37	56	17.7	155	10	US-09-981-876-164	Sequence 164, App
38	56	17.7	155	10	US-09-981-876-164	Sequence 225, App
39	56	17.7	155	11	US-09-148-545-225	Sequence 164, App
40	56	17.7	155	11	US-09-148-545-225	Sequence 225, App
41	56	17.7	209	12	US-10-094-749-2490	Sequence 2490, Ap
42	56	17.7	699	9	US-09-759-359A-2	Sequence 2, Appl
43	56	17.7	699	12	US-10-207-973-2	Sequence 2, Appl
44	55.5	17.6	556	10	US-09-801-368-426	Sequence 426, App
45	55	17.4	49	10	US-09-764-846-228	Sequence 228, App

#### ALIGNMENTS

RESULT 1  
US-10-079-754A-7  
Sequence 7, Application US/10079754A  
; PUBLIC ID: US0020164625A1  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Grigor, Murray R.  
; APPLICANT: Molenaar, Adrian J.  
; APPLICANT: Davis, Stephen R.  
; TITLE OF INVENTION: Compositions Isolated from Bovine  
; FILE REFERENCE: 11000.1068  
; CURRENT APPLICATION NUMBER: US/10/079, 754A  
; PRIOR FILING DATE: 2002-02-19  
; PRIOR APPLICATION NUMBER: US 09/699, 146  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US 60,162,701  
; PRIOR FILING DATE: 1999-10-29  
; PRIOR APPLICATION NUMBER: US 09/644,190  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: US 60,150,330  
; PRIOR FILING DATE: 1999-08-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 7  
; LENGTH: 58  
; TYPE: PRT  
; ORGANISM: Bovine  
US-10-079-754A-7

Query Match 100.0%; Score 316; DB 14; Length 58;  
Best Local Similarity 100.0%; Pred. No. 3e-30;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 MKRIFVFVIMLILMIRADSSSEKRRKRRKHGRGYQQYQYQYRYPLNYPAYPP 58  
Db 1 MKRIFVFVIMLILMIRADSSSEKRRKRRKHGRGYQQYQYQYRYPLNYPAYPP 58

```
RESULT 2
US-10-079-754A-10
; Sequence 10, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Grigor, Matthew
; APPLICANT: Molesnaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions Isolated from Bovine
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079,754A
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/699,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60,162,701
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/644,190
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 60,150,330
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Bovine
US-10-079-754A-10

Query Match          100.0%; Score 316; DB 14; Length 58;
Best Local Similarity 100.0%; Pred. No. 3e-30;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MKIFIFVFMALILAMIRADSSSEKRRKRRKGGYFOQYQYQRYPLNYPAYPP 58
Db      1 MKIFIFVFMALILAMIRADSSSEKRRKRRKGGYFOQYQYQRYPLNYPAYPP 58

RESULT 3
US-10-079-754A-8
; Sequence 8, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molesnaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions Isolated from Bovine
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079,754A
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/699,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60,162,701
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/644,190
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60,150,330
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Bovine
US-10-079-754A-8

Query Match          99.7%; Score 315; DB 14; Length 58;
Best Local Similarity 98.3%; Pred. No. 3.9e-30;
Matches 57; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 MKIFIFVFMALILAMIRADSSSEKRRKRRKGGYFOQYQYQRYPLNYPAYPP 58
Db      1 MKIFIFVFMALILAMIRADSSSEKRRKRRKGGYFOQYQYQRYPLNYPAYPP 58

RESULT 4
US-10-079-754A-12
; Sequence 12, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Grigor, Matthew
; APPLICANT: Molesnaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions Isolated from Bovine
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079,754A
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/699,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60,162,701
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/644,190
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 60,150,330
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Bovine
US-10-079-754A-12

Query Match          53.8%; Score 170; DB 14; Length 59;
Best Local Similarity 100.0%; Pred. No. 5.6e-13;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MKIFIFVFMALILAMIRADSSSEKRRKRRKHH 34
Db      1 MKIFIFVFMALILAMIRADSSSEKRRKRRKHH 34

RESULT 5
US-10-079-754A-9
; Sequence 9, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molesnaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions Isolated from Bovine
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079,754A
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/699,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60,162,701
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/644,190
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 60,150,330
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Bovine
US-10-079-754A-9
```

US-10-079-754A-9

Query Match	53.5%;	Score 169;	DB 14;	Length 70;
Best Local Similarity	97.1%;	Pred. No. 8.8e-13;		
Matches 33; Conservative	1;	Mismatches 0;	Indels	

Qy	1	MKIFIFVFIMALLAMIRADSSSEKKHRRKKCH	34
		:	
Db	1	MKIFIFIFIMALLAMIRADSSSEKKHRRKKCH	34

RESULT 6  
US-10-079-754A-11

```

Sequence 11, Application US/1,007,975A4
Publication No. US20020164625A1
GENERAL INFORMATION:
APPLICANT: Glenn, Matthew
APPLICANT: Grigor, Murray R.
APPLICANT: Molenaar, Adrian J.
APPLICANT: Davis, Stephen R.
TITLE OF INVENTION: Mammary Gland and Methods for Bovine
FILE REFERENCE: 11000.1068
CURRENT APPLICATION NUMBER: US/10/079, 754A
CURRENT FILING DATE: 2002-02-19
PRIOR APPLICATION NUMBER: US 09/659, 146
PRIOR FILING DATE: 2000-10-27
PRIOR APPLICATION NUMBER: US 60,162,701
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: US 09/644,190
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: US 60,150,330
PRIOR FILING DATE: 1999-08-23
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 21
TYPE: PR1
ORGANISM: Bovine
US-10-079-754A-11

```

Query Match	35.4%	Score 112	DB 14	Length 21
Best Local Similarity	94.7%	Pred. No. 1.3e-06		
Matches 18	Conservative 1	Mismatches 0	Indels 0	Gaps 0

Qy	38	FQOYQPYQRYPPLNPPAYP	56
		:	
Db	1	FQOYQPYERYPLNPPAYP	19

RESULT 7  
US-09-917-340-19

```

? Sequence 19, US/09/917,340
? Patent No. US20020090369A1
? GENERAL INFORMATION:
? APPLICANT: Murphy, Christopher J.
? APPLICANT: McNulty, Jonathan F.
? APPLICANT: Reid, Ted W.
? TITLE OF INVENTION: Transplant Media
? FILE REFERENCE: TPLANT-06468
? CURRENT APPLICATION NUMBER: US/09/917,340
? CURRENT FILING DATE: 2001-07-29
? PRIOR APPLICATION NUMBER: 60/221,632
? PRIOR FILING DATE: 2000-07-28
? PRIOR APPLICATION NUMBER: 60/249,602
? PRIOR FILING DATE: 2000-11-17
? PRIOR APPLICATION NUMBER: 60/290,932
? PRIOR FILING DATE: 2001-05-15
? NUMBER OF SEQ ID NOS: 96
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 19
? LENGTH: 51
? TYPE: PRT

```

US-09-917-340-19

Query Match	30.1%	Score 95;	DB 9;	Length 51;
Best Local Similarity	46.8%;	Pred. No. 0.00035;		
Matches 22;	Conservative	6;	Mismatches 13;	Indels

[illegible]

RESULT 8  
US-09-992-600A-6

sequence 6, Application US/0992600A  
Publication No. US20030027161A1  
GENERAL INFORMATION:  
APPLICANT: Benjamin, Stephane  
APPLICANT: Tanaka, Hiroki  
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
FILE REFERENCE: 91.USA.DIV  
CURRENT APPLICATION NUMBER: US/09/992,600A  
CURRENT FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 09/924,340  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: PCT/IB01/01715  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/305,456  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/302,277  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/298,698  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/293,574  
PRIOR FILING DATE: 2001-05-25  
NUMBER OF SEQ. ID NOS: 114  
SOFTWARE: JPatent  
SEQ ID NO 6  
LENGTH: 78  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: 1..19  
US-09-992-600A-6

Query Match	29.9%	Score 94.5	DB 11	Length 78
Best Local Similarity	32.7%	Pred. No. 0.00064		
Matches 18	Conservative 13	Mismatches 17	Indels 7	Gaps 1

```
QY      1 MKIFIFVIMALLIAMIADSDSEKRRKRKKHHGYYFOOYPDYORYPPLNPPAY 55
       |||:::||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      1 MKFVFVALVLAMTSMISADSHSKRHGGYRKFF-----EKHSHYITLLPLF 48
```

RESULT 9  
US-09-924-340-6

Sequence 6, Application US/09924340  
Publication No. US20030027248A1  
GENERAL INFORMATION:  
APPLICANT: Bejainin, Stephane  
APPLICANT: Tanaka, Hiroaki  
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
FILE REFERENCE: 91, US2, REG  
CURRENT APPLICATION NUMBER: US/09/924,340  
CURRENT FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/305,456  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/302,277  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/298,698  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/293,574

;; PRIOR FILING DATE: 2001-05-25  
;; NUMBER OF SEQ ID NOS: 112  
;; SOFTWARE: Jpatent  
;; SEQ ID NO 6  
;; LENGTH: 78  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SIGNAL  
;; LOCATION: 1..19  
US-09-924-340-6

Query Match 29.9%; Score 94.5; DB 11; Length 78;  
Best Local Similarity 32.7%; Pred. No. 0.00064;  
Matches 18; Conservative 13; Mismatches 17; Indels 7; Gaps 1;

QY 1 MKTFFVFMALIAMIRADSSSEKRRKRRKGGYFOQYQYRYPLNYPAY 55  
DB 1 MKFFVFLVIALMISMISADSHKRRHGGYRRKFH-----EKHSYHITLLPLF 48

RESULT 10  
US-09-992-095B-6  
;; Sequence 6, Application US/09992095B  
;; Publication No. US20030157485A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Benjamin, Stephanie  
;; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
;; FILE REFERENCE: 91.US5.DIV  
;; CURRENT APPLICATION NUMBER: US/09/992,095B  
;; PRIOR FILING DATE: 2003-02-20  
;; PRIOR APPLICATION NUMBER: US 09/924,340  
;; PRIOR FILING DATE: 2001-08-06  
;; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
;; PRIOR FILING DATE: 2001-08-06  
;; PRIOR APPLICATION NUMBER: US 60/305,456  
;; PRIOR FILING DATE: 2001-07-13  
;; PRIOR APPLICATION NUMBER: US 60/302,277  
;; PRIOR FILING DATE: 2001-06-29  
;; PRIOR APPLICATION NUMBER: US 60/298,698  
;; PRIOR FILING DATE: 2001-06-15  
;; PRIOR APPLICATION NUMBER: US 60/293,574  
;; PRIOR FILING DATE: 2001-05-25  
;; NUMBER OF SEQ ID NOS: 112  
;; SOFTWARE: Jpatent  
;; SEQ ID NO 6  
;; LENGTH: 78  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SIGNAL  
;; LOCATION: 1..19  
US-09-992-095B-6

Query Match 29.9%; Score 94.5; DB 12; Length 78;  
Best Local Similarity 32.7%; Pred. No. 0.00064;  
Matches 18; Conservative 13; Mismatches 17; Indels 7; Gaps 1;

QY 1 MKTFFVFMALIAMIRADSSSEKRRKRRKGGYFOQYQYRYPLNYPAY 55  
DB 1 MKFFVFLVIALMISMISADSHKRRHGGYRRKFH-----EKHSYHITLLPLF 48

RESULT 11  
US-10-154-678-6  
;; Sequence 6, Application US/10154678  
;; Publication No. US20030162186A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Benjamin, Stephanie  
;; APPLICANT: Tanaka, Hiroaki  
;; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
;; FILE REFERENCE: 182.US1.REG

;; CURRENT APPLICATION NUMBER: US/10/154,678  
;; CURRENT FILING DATE: 2002-10-15  
;; PRIOR APPLICATION NUMBER: US 09/924,340  
;; PRIOR FILING DATE: 2001-08-06  
;; PRIOR APPLICATION NUMBER: US 60/305,456  
;; PRIOR FILING DATE: 2001-07-13  
;; PRIOR APPLICATION NUMBER: US 60/302,277  
;; PRIOR FILING DATE: 2001-06-29  
;; PRIOR APPLICATION NUMBER: US 60/298,698  
;; PRIOR FILING DATE: 2001-06-15  
;; PRIOR APPLICATION NUMBER: US 60/293,574  
;; PRIOR FILING DATE: 2001-05-25  
;; NUMBER OF SEQ ID NOS: 112  
;; SOFTWARE: Jpatent  
;; SEQ ID NO 6  
;; LENGTH: 78  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SIGNAL  
;; LOCATION: -19..-1  
US-10-154-678-6

Query Match 29.9%; Score 94.5; DB 12; Length 78;  
Best Local Similarity 32.7%; Pred. No. 0.00064;  
Matches 18; Conservative 13; Mismatches 17; Indels 7; Gaps 1;

QY 1 MKTFFVFMALIAMIRADSSSEKRRKRRKGGYFOQYQYRYPLNYPAY 55  
DB 1 MKFFVFLVIALMISMISADSHKRRHGGYRRKFH-----EKHSYHITLLPLF 48

RESULT 12  
US-09-999-570-6  
;; Sequence 6, Application US/09999570  
;; Publication No. US20030170628A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Benjamin, Stephanie  
;; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
;; FILE REFERENCE: G-091US08DIY  
;; CURRENT APPLICATION NUMBER: US/09/999,570  
;; CURRENT FILING DATE: 2001-06-14  
;; PRIOR APPLICATION NUMBER: US 09/924,340  
;; PRIOR FILING DATE: 2001-08-06  
;; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
;; PRIOR FILING DATE: 2001-08-06  
;; PRIOR APPLICATION NUMBER: US 60/305,456  
;; PRIOR FILING DATE: 2001-07-13  
;; PRIOR APPLICATION NUMBER: US 60/302,277  
;; PRIOR FILING DATE: 2001-06-29  
;; PRIOR APPLICATION NUMBER: US 60/298,698  
;; PRIOR FILING DATE: 2001-06-15  
;; PRIOR APPLICATION NUMBER: US 60/293,574  
;; PRIOR FILING DATE: 2001-05-25  
;; NUMBER OF SEQ ID NOS: 112  
;; SOFTWARE: Jpatent  
;; SEQ ID NO 6  
;; LENGTH: 78  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SIGNAL  
;; LOCATION: 1..19  
US-09-999-570-6

Query Match 29.9%; Score 94.5; DB 12; Length 78;  
Best Local Similarity 32.7%; Pred. No. 0.00064;  
Matches 18; Conservative 13; Mismatches 17; Indels 7; Gaps 1;

QY 1 MKTFFVFMALIAMIRADSSSEKRRKRRKGGYFOQYQYRYPLNYPAY 55  
DB 1 MKFFVFLVIALMISMISADSHKRRHGGYRRKFH-----EKHSYHITLLPLF 48

```
RESULT 13
US-10-000-489-6
; Sequence 6, Application US/10000489
; Publication No. US20030092011A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US.6.DIV
; CURRENT APPLICATION NUMBER: US/10/000,489
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: Jpatent
; SEQ ID NO 6
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: 1..19
US-10-000-489-6

Query Match          29.9%; Score 94.5; DB 15; Length 78;
Best Local Similarity 32.7%; Pred. No. 0.00064;
Matches 18; Conservative 13; Mismatches 17; Indels 7; Gaps 1;

QY 1 MKLFFVIMLILAMIRADSEKRRKHHGFGYQYQYQYRPLNPPAY 55
DB 1 MKFFVFLVLMISMISADSHKHHGRRKFF-----EKHSHYITLLPLF 48

RESULT 14
US-10-000-986-6
; Sequence 6, Application US/10000986
; Publication No. US20030096247A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US.9.DIV
; CURRENT APPLICATION NUMBER: US/10/000,986
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: Jpatent
; SEQ ID NO 6
; LENGTH: 78
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: 1..19
US-10-000-986-6

Query Match          29.9%; Score 94.5; DB 15; Length 78;
Best Local Similarity 32.7%; Pred. No. 0.00064;
Matches 18; Conservative 13; Mismatches 17; Indels 7; Gaps 1;

QY 1 MKLFFVIMLILAMIRADSEKRRKHHGFGYQYQYQYRPLNPPAY 55
DB 1 MKFFVFLVLMISMISADSHKHHGRRKFF-----EKHSHYITLLPLF 48

RESULT 15
US-10-142-077-33
; Sequence 33, Application US/10142077
; Publication No. US2003007619A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Monto H.
; APPLICANT: delia-Cioppa, Guy R.
; APPLICANT: Erwin, Robert L.
; APPLICANT: McGee, David R.
; TITLE OF INVENTION: METHOD OF DETERMINING THE PRESENCE OF A
; TITLE OF INVENTION: TRAIT IN A PLANT BY TRANSFECTING A NUCLEIC ACID SEQUENCE OF
; TITLE OF INVENTION: A NON-PLANT DONOR INTO A HOST PLANT IN AN ANTI-SENSE
; TITLE OF INVENTION: ORIENTATION
; FILE REFERENCE: 008010137US06
; CURRENT APPLICATION NUMBER: US/10/142,077
; PRIOR FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US/09/359,297
; PRIOR FILING DATE: 1999-07-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-142-077-33

Query Match          19.9%; Score 63; DB 15; Length 103;
Best Local Similarity 54.2%; Pred. No. 4.6;
Matches 13; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 35 RGYFOGYQYQYRPLNPPAYPP 58
DB 67 QGYFOGYQYQYRPLNPPAYPP 90
```

Search completed: December 13, 2003, 03:14:20  
Job time : 276 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 13, 2003, 03:08:45 / Search time 2424 Seconds

(without alignments)  
976.861 Million cell updates/sec

Title: US-10-079-754a-10

Perfect score: 316  
Sequence: 1 MKIFIFVFIMLILMIRAD.....QQYQYQYRPLNYPPEYFP 58

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1h  
-O=/sgn2.1/USPTO.spool/US1007975/runat\_12122003\_145752\_13012/app.query.fasta\_1.199  
-DB=genemb1 -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCLIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10079754 @CGN 1.1 3508 @runat\_12122003\_145752\_13012 -NCFU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1:\*

1:	gb_ba:*
2:	gb_hcg:*
3:	gb_in:*
4:	gb_om:*
5:	gb_ov:*
6:	gb_pac:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_sts:*
12:	gb_ay:*
13:	gb_un:*
14:	gb_vl:*
15:	em_ba:*
16:	em_fun:*
17:	em_hum:*
18:	em_in:*
19:	em_mu:*
20:	em_om:*
21:	em_or:*
22:	em_ov:*
23:	em_pac:*
24:	em_ph:*
25:	em_pl:*
26:	em_ro:*
27:	em_sts:*
28:	em_un:*

29: em\_vl:\*

30: em\_hcg\_hum:\*

31: em\_hcg\_inv:\*

32: em\_hcg\_other:\*

33: em\_hcg\_mus:\*

34: em\_hcg\_pln:\*

35: em\_hcg\_rod:\*

36: em\_hcg\_mam:\*

37: em\_hcg\_vrt:\*

38: em\_ay:\*

39: em\_hcgo\_hum:\*

40: em\_hcgo\_mus:\*

41: em\_hcgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	146	46.2	195534	2 AC134173	AC134173 Bos tauru
2	123.5	39.1	486	4 AY154893	AY154893 Bos tauru
3	110	34.8	542	9 HUMSTRNA	M18371 Human stach
4	110	34.8	552	9 HUMSTRTHA	M18078 Human stach
5	110	34.8	1584	9 AKO92678	AKO92678 Homo sapi
6	102	32.3	480	9 H0WHISIX	M26664 Human hist
7	102	32.3	566	9 BC017835	BC017835 Homo sapi
8	99	31.3	357	6 BD071675	BD071675 Secreted
9	97	30.7	516	6 BD071674	BD071674 Secreted
10	95	30.1	491	9 HUMBRHRA	M18372 Human hist
11	95	30.1	524	9 HUMHIS2X	M26665 Human hist
12	95	30.1	558	9 BC009791	BC009791 Homo sapi
13	94.5	29.9	438	6 AX588147	AX588147 Sequence
14	94.5	29.9	438	6 AX588483	AX588483 Sequence
15	94.5	29.9	438	6 AX616321	AX616321 Sequence
16	82	25.9	176	6 BD071751	BD071751 Secreted
17	82	25.9	203	6 BD071749	BD071749 Secreted
18	80	25.3	223069	2 AC094956	AC094956 Rattus no
C 19	79.5	25.2	178027	2 AC132041	AC132041 Rattus no
C 20	77	24.4	201002	2 BX088549	BX088549 Dantio rer
C 21	76	24.1	216887	2 AC113676	AC113676 Rattus no
C 22	76	24.1	226674	2 AC131171	AC131171 Rattus no
C 23	76	24.1	246890	2 AC126807	AC126807 Mus muscu
24	75.5	23.9	617	3 TENSPE23A	M92928 Tenexrio mo
25	75.5	23.9	38692	3 AC116919	AC116919 Dictyoste
C 26	75.5	23.9	63489	8 AC023064	AC023064 Arabidops
27	75	23.7	756	3 AY119164	AY119164 Drosophill
28	75	23.7	78507	2 AC019704	AC019704 Drosophill
29	75	23.7	101647	8 AC004665	AC004665 Arabidops
30	75	23.7	132544	1 AF521085	AF521085 Streptomy
31	75	23.7	169210	3 AC092226	AC092226 Drosophill
32	75	23.7	176306	2 AC007186	AC007186 Drosophill
33	75	23.7	258024	2 AC084046	AC084046 Trypanoso
34	75	23.7	289460	3 AE003630	AE003630 Drosophill
35	74.5	23.6	91440	2 OSJND0044	OSJND0044 Arabidops
36	74.5	23.6	138713	8 OSJND0190	OSJND0190 Arabidops
C 37	74	23.4	173594	2 AC135597	AC135597 Oryza sat
C 38	74	23.4	191261	2 AC074304	AC074304 Rattus no
C 39	74	23.4	196729	2 AC117955	AC117955 Rattus no
40	74	23.4	202915	10 AL928829	AL928829 Mouse DNA
C 41	74	23.4	211300	2 AC131666	AC131666 Mus muscu
42	74	23.4	249047	2 AC120964	AC120964 Rattus no
C 43	74	23.4	316311	2 AC125917	AC125917 Rattus no
44	73	23.1	210800	2 AC127824	AC127824 Rattus no
45	73	23.1	226542	5 AL805947	AL805947 Zebrafish

RESULT 1

## ALIGNMENTS

AC134173/c  
LOCUS AC134173 195534 bp DNA linear HTG 27-SEP-2002  
DEFINITION Bos taurus clone RP42-254113, WORKING DRAFT SEQUENCE, 26 unordered  
pieces.  
AC134173  
VERSION AC134173.1 GI:23306007  
KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 195534)  
REFERENCE  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
Alzubrooke, S.L., Amaralunga, H.C., Are, J.R., Ayale, M., Banks, T.,  
Barberia, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dedertich, D.A.,  
DeLaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earmhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P.,  
Gabriel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorelli, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, J., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,  
Homs, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karleson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, U., Liu, W., Louised, H.,  
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawliny, E., McLeod, M.P., Meador, K., Morgan, M., Morris, S.,  
Mosier, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogund, M., Okunolu, G.,  
Ogunyeye, N., Oyale, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Scheerer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, U., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
TITLE  
JOURNAL Unpublished  
REFERENCE  
AUTHORS 2 (bases 1 to 195534)  
JOURNAL Worley, K.C.  
TITLE Direct Submission  
REFERENCE  
AUTHORS Submitted (24-SEP-2002) Human Genome Sequencing Center, Department  
JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 195534)  
REFERENCE  
AUTHORS Worley, K.C.  
TITLE Direct Submission  
JOURNAL Submitted (27-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information

Center project name: VUAE  
Center clone name: RP42-254113  
----- Summary Statistics  
Sequencing vector: M13;  
Chemistry: Dye-terminator Big Dye; 8% of reads  
Chemistry: Dye-terminator Big Dye; 18% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 185623 bases at least Q40  
Consensus quality: 190257 bases at least Q30  
Consensus quality: 192456 bases at least Q20  
Estimated insert size: 204493; sum-of-coverage estimation  
Quality coverage: 4x in Q20 bases; sum-of-coverage estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
2358: contig of 2358 bp in length  
2359: gap of unknown length  
2459: contig of 3161 bp in length  
5620: gap of unknown length  
5720: contig of 2591 bp in length  
8310: gap of unknown length  
8411: contig of 2592 bp in length  
11003: gap of unknown length  
11103: contig of 3284 bp in length  
14387: gap of unknown length  
14487: contig of 2663 bp in length  
17487: gap of unknown length  
1750: contig of 4165 bp in length  
21415: gap of unknown length  
21515: contig of 3657 bp in length  
25172: gap of unknown length  
25272: contig of 5374 bp in length  
30646: gap of unknown length  
30746: contig of 5437 bp in length  
36182: gap of unknown length  
36283: contig of 5674 bp in length  
41957: gap of unknown length  
42056: contig of 6038 bp in length  
42057: gap of unknown length  
48095: contig of 5020 bp in length  
48195: gap of unknown length  
53215: contig of 5535 bp in length  
53315: gap of unknown length  
58849: contig of 7195 bp in length  
58950: gap of unknown length  
66145: contig of 6789 bp in length  
66245: gap of unknown length  
73034: contig of 6897 bp in length  
73134: gap of unknown length  
80031: contig of 7631 bp in length  
80131: gap of unknown length  
8761: contig of 7865 bp in length  
87762: gap of unknown length  
95726: contig of 9228 bp in length  
95827: gap of unknown length  
105055: contig of 11562 bp in length  
105154: gap of unknown length  
116716: contig of 11477 bp in length  
116717: gap of unknown length  
128294: contig of 9747 bp in length  
128394: gap of unknown length  
138141: contig of 12447 bp in length  
138241: gap of unknown length  
150687: contig of 15677 bp in length  
150787: gap of unknown length  
165414: contig of 15677 bp in length  
165415: gap of unknown length

Score:	123.50	Matches:	32
Percent Similarity:	62.71%	Conservative:	19
Best Local Similarity:	54.24%	Mismatches:	15
Query Match:	39.08%	Indels:	3
	4	Gaps:	3
US-10-079-754A-10 (1-58) x AY154893 (1-486)			
QY	1 MethyallephelelePhevalPheIIMetalauleleualamelleargNlaasp	20	
Db	1 ATGAAGATCTTTTCTTGCCCTTATATGCTCTCAATGTTGCATGATTAAGCTGAT	60	
QY	21 SerSerGluGluysaGhIaargLyArgLyshIshIaAGGlyTYrPheGluGln	40	
Db	61 TCATCTTAAGAGGAA---CACCGGCTTAGATTAAACCTAGATTATTAACCAAAACAG	117	
QY	41 TyGlnProTYr---GlnArgTYrProLeuasnTYrProalaTYrProPhePro	58	
Db	118 CAAGTGGATATATCTCTAGTATTCACGA---TATCATATTCATATTCATATCA	171	
RESULT 3			
HUMSTRNA	542 bp	mRNA	Linear PRI 13-JAN-1995
LOCUS	Human statherin mRNA, complete cds.		
DEFINITION	M18371		
ACCESSION	M18371.1 GI:338610		
VERSION	statherin.		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 542)		
AUTHORS	Dickinson,D.P., Ridall,A.L. and Levine,M.J.		
TITLE	Human submandibular gland statherin and basic histidine-rich peptide are encoded by highly abundant mRNA's derived from a common ancestral sequence		
JOURNAL	Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)		
MEDLINE	88106506		
PUBMED	Original source text: Human female submandibular gland, cDNA to mRNA, clone pBRMS5F98.2.		
COMMENT	Draft entry and computer readable copy of sequence [1] kindly provided by D.P. Dickinson 21-MAR-1988.		
FEATURES	Location/Qualifiers		
source	1..542		
gene	/organism="Homo sapiens"		
CDS	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/map="4q11-q13"		
	1..542		
	/gene="STATH"		
	57..245		
	/gene="STATH"		
	/note="statherin precursor"		
	/codon_start=1		
	/protein_id="AA60600.1"		
	/db_xref="GI:338611"		
	/db_xref="GDB:G00-120-391"		
	/translation="MKRLVPAFILALVMSIGASSSEKFLRRIGRGYGVGPQPVH		
	EQPLVPQVQQTTF"		
	57..113		
sig_peptide	/gene="STATH"		
	/note="statherin signal peptide"		
mat_peptide	114..242		
	/gene="STATH"		
	/product="statherin"		
BASE COUNT	170 a 117 c 76 g 179 t		
ORIGIN	262 bp upstream of PstI site; chromosome 4q11-q13.		
Alignment Scores:			
Pred. No.:	0.000312	Length:	542
Score:	110.00	Matches:	27
Percent Similarity:	61.82%	Conservative:	7

Best Local Similarity: 49.09% Mismatches: 19  
 Query Match: 34.81% Indels: 2  
 DB: 9 Gaps: 2

US-10-079-754a-10 (1-58) x HUMSTATHA (1-542)

Qy 1 MetLysIlePheIleValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaasp 20

Db 57 ATGAAGTTCCTTGCTTGGCTTCATCTTGCTCTCATGTTCCATGATGAGCTGAT 116

Qy 21 SerSerGIuGIuLysArgHisArgLysArgLysHisHisArgGIyTYrPheGlnGln 40

Db 117 TCATCTGAAGAGAAATTTTGGCTTGAATGGAATTCGCTTATGCGTAT---GGCCCT 173

Qy 41 TYrGlnProTYrGlnArgTYrProLeuAanTYrProProAlaTYr 55

Db 174 TATCAGCCAGTTCAGAACACCACTA---TACCACACCACTATC 215

RESULT 4 HUMSTATHA 552 bp mRNA linear PRI 13-JAN-1995

LOCUS Human steatherin mRNA, complete cds.

DEFINITION M18078.1 GI:338507

VERSION regulatory protein; steatherin.

KEYWORDS Homo sapiens (human)

SOURCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 552)

AUTHORS Sabatini,L.M., Carlsson,L.R., Johnson,G.W. and Azen,E.A.

TITLE cDNA cloning and chromosomal localization (4q11-13) of a gene for

steatherin, a regulator of calcium in salivary

JOURNAL Am. J. Hum. Genet. 41 (6), 1048-1060 (1987)

MEDLINE 88074310

PUBMED 3502720

COMMENT Original source text: Human parotid gland, cDNA to mRNA, clone

H772B.

Draft entry and computer-readable sequence [1] kindly submitted by

L.Sabatini 19-JAN-1988.

FEATURES

source

1..552

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/map="4q11-q13"

1..552

/gene="STATH"

73..261

/note="STATH"

/note="steatherin protein"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/gene="STATH"

/note="steatherin signal peptide"

/codon\_start=1

/protein\_id="AAA60594.1"

/db\_xref="GI:338508"

/translation="MKPLVFAFIALAVSMIGADSEKFLRIGRGYGYQYQYVP

EQLYPOQYPOQYQYTF"

73..129

/

Pred. No.: 0.00102 Length: 1584  
 Score: 110.00 Matches: 27  
 Percent Similarity: 61.82% Conservative: 7  
 Best Local Similarity: 49.09% Mismatches: 19  
 Query Match: 34.81% Indels: 2  
 DB: 9 Gaps: 2

US-10-079-754A-10 (1-58) x AK092678 (1-1584)

QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuAlaMetIleArgAlaAsp 20  
 ||||| ..... ||||| ..... ||||| ..... ||||| ..... ||||| .....  
 Db 1097 ATGAAGTTCCTTCTCTTCCTTCATCTTGCTTCATGCTTCATGATTCGAGCTGAT 1156

QY 21 SerSerGluGluValArgHisArgLysArgLysHisHisArgGlyTyrPheGlnGln 40  
 ||||| ..... ||||| ..... ||||| ..... ||||| ..... ||||| .....  
 Db 1157 TCATCTGAAGAAGAAATTTTGGCTAGAAATGGAAGATTGGTTATGGGTAT---GGCCT 1213

QY 41 TyrGlnProTyrGlnArgTyrProLeuAnTyProProAlaTyr 55  
 ||||| ..... ||||| ..... ||||| ..... ||||| ..... ||||| .....  
 Db 1214 TATCAGCAGTTCAGAACACCACTA--TACCCACACCACTAC 1255

RESULT 6  
 HUMHIS1X  
 LOCUS HUMHIS1X 480 bp mRNA linear PRI 31-DEC-1994  
 DEFINITION Human histatin 1 (HIS1) mRNA, complete cds.  
 ACCESSION M26664  
 VERSION M26664.1 GI:292143  
 KEYWORDS histatin 1.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Sabatini, L.M. and Azen, E.A.  
 TITLE Histatins, a family of salivary histidine-rich proteins, are encoded by at least two loci (HIS1 and HIS2)  
 JOURNAL Biochem. Biophys. Res. Commun. 160 (2), 495-502 (1989)  
 MEDLINE 89246491  
 PUBMED 2719677

COMMENT Original source text: Homo sapiens parotid gland cDNA to mRNA.  
 FEATURES  
 source  
 1..480  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /tissue\_type="parotid gland"  
 1..480  
 /gene="HIS1"  
 10..183  
 /gene="HIS1"  
 /codon\_start=1  
 /product="histatin 1"  
 /protein\_id="AA58645.1"  
 /db\_xref="GI:292144"  
 /translation="MKFFVFLVALMISMSADSHKRRHGRKRRPEKHSRREP  
 FYGDGSMVLYDN"

polyA\_signal  
 462..467  
 /gene="HIS1"  
 BASE COUNT 152 a 85 c 75 g 168 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.00254 Length: 480  
 Score: 102.00 Matches: 20  
 Percent Similarity: 62.26% Conservative: 13  
 Best Local Similarity: 37.74% Mismatches: 18  
 Query Match: 32.28% Indels: 2  
 DB: 9 Gaps: 1

US-10-079-754A-10 (1-58) x HUMHIS1X (1-480)  
 QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuAlaMetIleArgAlaAsp 20  
 ||||| ..... ||||| ..... ||||| ..... ||||| ..... ||||| .....  
 ||||| ..... ||||| ..... ||||| ..... ||||| ..... ||||| .....

Db 10 ATGAAGTTCCTTCTCTTCCTTCATGCTTCATGATTCGATTCAGCGCTGAT 69  
 QY 21 SerSerGluGluValArgHisArgLysArgLysHisHisArgGlyTyr-----Phe 38  
 ||||| ..... ||||| ..... ||||| ..... ||||| ..... ||||| .....  
 Db 70 TCATCTGAAGAAGACATCATGCTGATAGAAATTCATGAAGAAGCATCATTCACAT 129  
 QY 39 GlnGlnTyrGlnProTyrGlnArgTyrProLeuAnTyProProAlaTyr 51  
 ||||| ..... ||||| ..... ||||| ..... ||||| ..... ||||| .....  
 Db 130 CGAAGATTCATCTTTATGGGGACTATGATCAATATAT 168

RESULT 7  
 LOCUS BC017835 566 bp mRNA linear PRI 06-DEC-2001  
 DEFINITION Homo sapiens, histatin 1, clone MGC:22502 IMAGE:4289874, mRNA, complete cds.  
 ACCESSION BC017835  
 VERSION BC017835.1 GI:17389614  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbe-remail.nih.gov](mailto:cgapbe-remail.nih.gov)

Tissue Procurement: CLONTECH  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www.shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcdeppaxil.stanford.edu](mailto:mcdeppaxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IPAL Plate: 36 Row: A Column: 21  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504528.  
 FEATURES  
 source  
 1..566  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /db\_xref="LocusID:3346"  
 /db\_xref="taxon:9606"  
 /clone="MGC:22502 IMAGE:4289874"  
 /tissue\_type="Skeletal Muscle"  
 /clone\_id="NTH MGC\_81"  
 /lab\_host="DH10B"  
 /note="Vector: pDNR-LIB"  
 68..241  
 /codon\_start=1  
 /product="histatin 1"  
 /protein\_id="AAH17835.1"  
 /db\_xref="GI:17389615"  
 /translation="MKFFVFLVALMISMSADSHKRRHGRKRRPEKHSRREP  
 FYGDGSMVLYDN"

CDS  
 68..241  
 /note="Vector: pDNR-LIB"

BASE COUNT 192 a 102 c 84 g 168 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.00305 Length: 566  
 Score: 102.00 Matches: 20  
 Percent Similarity: 62.26% Conservative: 13  
 Best Local Similarity: 37.74% Mismatches: 18



SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS 1 (bases 1 to 491)  
TITLE Human submandibular gland secretin and basic histidine-rich  
peptide are encoded by highly abundant mRNA's derived from a common  
ancestral sequence.  
JOURNAL Biochem. Res. Commun. 149 (2), 784-790 (1987)  
MEDLINE 88106506  
PUBMED 3426601  
COMMENT Original source text: Human female submandibular gland, cDNA to  
mRNA, clone pBRHSM51B.1.  
Draft entry and computer readable copy of sequence [1] kindly  
provided by D.P. Dickinson 21-MAR-1988.  
FEATURES  
source  
1..491  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="3q21-qter"  
1..491  
/gene="HRG"  
38..193  
/gene="HRG"  
/note="precursor"  
/codon\_start=1  
/product="basic histidine-rich protein"  
/protein\_id="AA51830.1"  
/db\_xref="GI:179466"  
/db\_xref="GDB:G00-120-055"  
/translation="MKFVFALILALMLMTGADSHAKRHGKPKFKHSHRGYR  
SNLYLDN"  
38..94  
/gene="HRG"  
/note="basic histidine-rich protein signal peptide"  
95..190  
/gene="HRG"  
/product="basic histidine-rich protein"  
BASE COUNT 151 a 90 c 87 g 163 t  
ORIGIN Unreported.  
Alignment Scores:  
Pred. No.: 0.0183 Length: 491  
Score: 95.00 Matches: 22  
Percent Similarity: 59.57% Conservative: 6  
Best Local Similarity: 46.81% Mismatches: 13  
Query Match: 30.06% Indels: 6  
DB: 9 Gaps: 2  
US-10-079-754A-10 (1-58) x HUMHIS2X (1-491)  
QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaasp 20  
|||||  
Db 38 ATGAAGTTTTCCTTTCCTTAATCTTGCCTCCTCATGCTTCATCATGAGCTGAT 97  
QY 21 SerSerGluGluLysArgHis--ArgLysArgLysLysHis-----His 34  
|||||  
Db 98 TCACATGCAAGAGACATCATGGGTATTAAGAAATTCATGAAAGCATCATTCACAT 157  
QY 35 ArgGlyTyrPheGlnGlnTyr 41  
|||||  
Db 158 CGAGGCTATAGATCAAAATTTAT 178  
RESULT 11  
LOCUS HUMHIS2X 524 bp mRNA linear PRI 31-DEC-1994  
DEFINITION Human histatin 2 (HIS2) mRNA, complete cds.  
ACCESSION M26665.1 GI:292145  
VERSION M26665.1  
KEYWORDS histatin 2.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS Sabatini, L.M. and Azen, E.A.  
TITLE Histatins, a family of salivary histidine-rich proteins, are  
encoded by at least two loci (HIS1 and HIS2)  
JOURNAL Biochem. Res. Commun. 160 (2), 495-502 (1989)  
MEDLINE 89246491  
PUBMED 2719677  
COMMENT Original source text: Homo sapiens parotid gland cDNA to mRNA.  
FEATURES  
source  
1..524  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="parotid gland"  
1..524  
/gene="HIS2"  
72..227  
/gene="HIS2"  
/codon\_start=1  
/product="histatin 2"  
/protein\_id="AA58646.1"  
/db\_xref="GI:292146"  
/translation="MKFVFALILALMLMTGADSHAKRHGKPKFKHSHRGYR  
SNLYLDN"  
506..511  
/gene="HIS2"  
BASE COUNT 157 a 100 c 87 g 180 t  
ORIGIN  
polyA\_signal  
Alignment Scores:  
Pred. No.: 0.0197 Length: 524  
Score: 95.00 Matches: 22  
Percent Similarity: 59.57% Conservative: 6  
Best Local Similarity: 46.81% Mismatches: 13  
Query Match: 30.06% Indels: 6  
DB: 9 Gaps: 2  
US-10-079-754A-10 (1-58) x HUMHIS2X (1-524)  
QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaasp 20  
|||||  
Db 72 ATGAAGTTTTCCTTTCCTTAATCTTGCCTCCTCATGCTTCATCATGAGCTGAT 131  
QY 21 SerSerGluGluLysArgHis--ArgLysArgLysLysHis-----His 34  
|||||  
Db 132 TCACATGCAAGAGACATCATGGGTATTAAGAAATTCATGAAAGCATCATTCACAT 191  
QY 35 ArgGlyTyrPheGlnGlnTyr 41  
|||||  
Db 192 CGAGGCTATAGATCAAAATTTAT 212  
RESULT 12  
LOCUS BC009791 558 bp mRNA linear PRI 12-JUL-2001  
DEFINITION Homo sapiens, histatin 3, clone MGC:13578 IMAGE:4293405, mRNA,  
complete cds.  
ACCESSION BC009791  
VERSION BC009791.1 GI:14602560  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS Straube, R.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUL-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK  
COMMENT  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: CLONTECH  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadan@systemsbio.org](mailto:amadan@systemsbio.org)  
Anup Madan, Rachel Dickhoff, Jessica Fahay, Stephanie Ford, Julia Greene, Mark Keltman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 19 Row: 0 Column: 1  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA GI: 4557652.  
Location/Qualifiers

FEATURES  
source  
1..558  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="LocusID:3347"  
/db\_xref="taxon:9606"  
/clone="MGC:13578 IMAGE:4293405"  
/tissue\_type="Skeletal Muscle"  
/clone\_lib="NIH MGC\_81"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
76..231  
/codon\_start=1  
/product="histatin 3"  
/protein\_id="AAH09791.1"  
/db\_xref="GI:14602561"  
/translation="MKFVFALILALMSMTGADSHAKRHGKFKHEKHSYHTL  
SNLYLDN"

CDS  
BASE COUNT 187 a 102 c 91 g 178 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.0211 Length: 558  
Score: 95.00 Matches: 22  
Percent Similarity: 59.57% Conservative: 6  
Best Local Similarity: 46.81% Mismatches: 13  
Query Match: 30.06% Indels: 6  
Gaps: 2

US-10-079-754A-10 (1-58) x BC009791 (1-558)

Qy 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaasp 20  
|||||  
Db 76 ATGAAGTTTGTGTTTGGCTTAATCTTGCGCTCTCATGCTTTCCATGACGAGCTGAT 135  
|||  
Qy 21 SerSerGluGluIuylsArgHls---ArgIylsArgIylsHls-----His 34  
|||  
Db 136 TCACATGCAAGAGACATCATGGGTATTAAGAAATTCATGAAAGCATCATTCACAT 195  
|||  
Qy 35 ArgGlyTyrPheGlnGlnTyr 41  
|||||  
Db 196 CGAGGCTATAGATCAATAT 216  
|||

RESULT 13  
AX588147  
LOCUS AX588147 438 bp DNA linear PAT 24-JAN-2003  
DEFINITION Sequence 22 from Patent WO02083898.  
ACCESSION AX588147  
VERSION AX588147.1 GI:27899822  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Benjamin, S., Tanaka, H., Dumas Milne Edwards, J.B., Jobert, S. and

TITLE Giordano, J.Y.  
JOURNAL Full-length human cdnas encoding potentially secreted proteins  
Patent: WO 02083898-A 22 24-OCT-2002;  
GENSET (FR)

FEATURES  
source  
Location/Qualifiers  
1..438  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
84..320  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD61372.1"  
/db\_xref="GI:27899823"  
/translation="MKFVFALILALMSMTGADSHAKRHGKFKHEKHSYHTL  
LPFRSSKSNANEKHNILYLCRIILAFSYT"  
84..140  
/note="Von Heijne matrix score 7.64030745849671 seq  
ALVIALMISMIS/DS"

CDS  
BASE COUNT 153 a 80 c 60 g 143 t 2 others  
ORIGIN

Alignment Scores:  
Pred. No.: 0.0186 Length: 438  
Score: 94.50 Matches: 18  
Percent Similarity: 56.36% Conservative: 13  
Best Local Similarity: 32.73% Mismatches: 17  
Query Match: 29.91% Indels: 7  
Gaps: 1

US-10-079-754A-10 (1-58) x AX588147 (1-438)

Qy 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaasp 20  
|||||  
Db 84 ATGAAGTTTGTGTTTGGCTTAATCTTGCGCTCTCATGCTTTCCATGATTCAGCGCTGAT 143  
|||  
Qy 21 SerSerGluGluIuylsArgHlsArgIylsArgIylsHlsArgIylsArgIyls 40  
|||||  
Db 144 TCACATGCAAGAGACATCATGGGTATTAAGAAATTCAT 185  
|||  
Qy 41 TyrGlnProTyrGlnIuylsArgIylsArgIylsHlsArgIylsArgIyls 55  
|||||  
Db 186 ---GAAAGCATCATTCATTCACATTCACACACTCTTCTT 227  
|||

RESULT 14  
AX588483  
LOCUS AX588483 438 bp DNA linear PAT 24-JAN-2003  
DEFINITION Sequence 358 from Patent WO02083898.  
ACCESSION AX588483  
VERSION AX588483.1 GI:27900156  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Benjamin, S., Tanaka, H., Dumas Milne Edwards, J.B., Jobert, S. and  
Giordano, J.Y.  
TITLE Full-length human cdnas encoding potentially secreted proteins  
Patent: WO 02083898-A 358 24-OCT-2002;  
GENSET (FR)

FEATURES  
source  
Location/Qualifiers  
1..438  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
84..320  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD61539.1"  
/db\_xref="GI:27900157"





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 13, 2003, 03:07:45 ; Search time 204 Seconds  
(without alignments)  
767,487 Million cell updates/sec

Title: US-10-079-754A-10  
Perfect score: 316  
Sequence: 1 MKRIFVFMILILAMIRAD.....QQYQYQYRPLNYPAYFPF 58

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756.seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgnr2.1/USPFTO.spool/US10079754/runat\_12122003\_145751\_13005/app\_query.fasta\_1.199  
-DB=N\_Geneseq\_19Jun03 -QFMT=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10079754.@CN\_1.1.0 @runat\_12122003\_145751\_13005 -NCPUL=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONELONG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03:\*  
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	316	100.0	267	25	ABX11394	Bovine mammary tis
2	316	100.0	505	25	ABX11388	Bovine histatin 1i
3	316	100.0	604	25	ABX11391	Bovine histatin 1i
4	315	99.7	585	25	ABX11389	Bovine histatin 1i
5	277	87.7	525	25	ABX11393	Bovine histatin 1i
6	170.5	54.0	869	25	ABX11390	Bovine histatin 1i
7	112	35.4	96	25	ABX11392	Bovine histatin 1i
8	110	34.8	552	21	AAA48964	Human streptococ
9	99	31.3	357	20	AAV89526	EST clone CP294.
10	97	30.7	516	20	AAV89525	EST clone CP289.
11	95	30.1	491	21	AAA48965	Human basic histid
12	95	30.1	857	22	AAH98658	Human EST-derived
13	94.5	29.9	438	25	ACC51062	Human Chimerin cod
14	94.5	29.9	438	25	ABZ36425	Human GENSER codin
15	94.5	29.9	438	25	ABZ36590	Human GENSER codin
16	86.5	27.4	457	24	ABQ99552	Human coding seque
17	82	25.9	176	20	AAV89602	EST clone CP92. H
18	82	25.9	203	20	AAV89600	EST clone CP41. H
19	78.5	24.8	2223	25	ABT18268	Aspergillus fumiga
20	78.5	24.8	2223	25	ABT18268	Aspergillus fumiga
21	78.5	24.8	2223	25	ABT20084	Aspergillus fumiga
22	78.5	24.8	2223	25	ABT20682	Aspergillus fumiga
23	78.5	24.8	4223	25	ABT17674	Aspergillus fumiga
24	78.5	24.8	4223	25	ABT19488	Aspergillus fumiga
25	75	23.7	696	23	ABL26229	Drosophila melanog
26	75	23.7	2696	23	ABL26228	Drosophila melanog
27	75	23.7	3677	23	ABL10500	Drosophila melanog
28	73	23.1	2121	21	AACT6665	Human ORFX ORF220
29	72.5	22.9	1353	23	ABL28449	Drosophila melanog
30	72.5	22.9	3597	23	ABL28448	Drosophila melanog
31	71	22.5	7388	23	ABV25625	Human prostate exp
32	70	22.2	3476	23	ABL27604	Drosophila melanog
33	69.5	22.0	517	24	ABK4762	cDNA encoding colo
34	69.5	22.0	1052	24	ABQ13620	Oligonucleotide fo
35	69.5	22.0	1052	24	ABQ13621	Oligonucleotide fo
36	69.5	22.0	2492	23	AAH86289	DNA encoding novel
37	69.5	22.0	2507	24	ABN60010	Novel human coding
38	69.5	22.0	2581	20	AAZ07215	Human lung tumour
39	69.5	22.0	2581	21	AACT9088	Human lung tumour
40	69.5	22.0	2581	23	AAZ23163	Human lung tumour
41	69.5	22.0	3085	23	ABL24014	Drosophila melanog
42	69.5	22.0	3097	24	ABO88094	Human osteoblast d
43	69.5	22.0	3097	24	ABO88095	Human osteoblast d
44	69.5	22.0	3899	24	AB199205	Mouse ischaemic co
45	68	21.5	315	21	AACT9238	Arabidopsis thalia

ALIGNMENTS

RESULT 1  
ID ABX11394 standard; cDNA; 267 BP.  
AC ABX11394;  
XX  
DT 01-MAY-2003 (first entry)  
XX  
DE Bovine mammary tissue cDNA clone.  
XX  
KW Cow; sis; histatin; bovine; salivary secretion; oral cavity;  
XX non-immune defence system; oral candidosis; gene therapy;  
KW microbial infection; fungal infection; dental caries; tartar;  
KW cystic fibrosis; systemic infection; Candida infection; mastitis;  
KW fungicide; antibacterial; mammary-gland.  
OS Bos taurus.

XX	US2002i64625-A1.
XX	
PD	07-NOV-2002.
XX	
PF	19-FEB-2002; 2002US-0079754.
XX	
PR	23-AUG-1999; 99US-150330P.
PR	29-OCT-1999; 99US-162701P.
PR	22-AUG-2000; 2000US-0644190.
PR	27-OCT-2000; 2000US-0699146.
PA	(GENE-) GENESIS RES & DEV CORP LTD.
PI	Glenn M, Grigor MR, Molenaar AJ, Davis SR;
PT	WPI: 2003-275306/27.
DR	
XX	
FT	New histatin polynucleotides and polypeptides expressed in bovine
PT	tissues, useful for treating bacterial or fungal infections in mammals,
PT	e.g. infections of the oral cavity, vagina, urethra, ear or skin, or
PT	systemic infections -
XX	
PS	Example 2; Page 15; 17pp; English.
XX	
CC	The invention discloses isolated histatin polynucleotides and
CC	polypeptides, which comprise sequences expressed in bovine. Histatins are
CC	histidine rich and specific to the salivary secretions. They are believed
CC	to function as part of the non-immune defence system, particularly in the
CC	oral cavity and have promise as therapeutic agents in humans with oral
CC	candidosis. Also disclosed are compositions comprising the histatin
CC	polypeptide or polynucleotide, and at least one component consisting of
CC	physiological or pharmaceutical carriers or immunostimulants. The
CC	histatin polynucleotide, polypeptide, or the cosmetic composition
CC	comprising the polypeptide, is useful for treating (e.g. gene therapy) a
CC	disorder in a mammal, particularly microbial or fungal infections. The
CC	histatin polynucleotide is also useful in genome mapping, physical
CC	mapping or in the positional cloning of genes. Specifically, the
CC	polypeptide or polynucleotide is useful for treating fungal or bacterial
CC	infections of the oral cavity (e.g. dental caries, plaque or tartar),
CC	vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis
CC	caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating
CC	systemic infections such as systemic Candida infection. The polypeptide
CC	or polynucleotide is also useful for veterinary applications e.g. for
CC	treating mastitis. The sequence presented is the bovine mammary tissue
CC	cDNA clone which was used to illustrate mRNA expression levels of
CC	mammary-gland specific sequences.
XX	
SQ	Sequence 267 BP; 89 A; 59 C; 43 G; 76 T; 0 other;
Alignment Scores:	
Pred. NO.:	2.52e-32 Length: 267
Score:	316.00 Matches: 58
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	25 Gaps: 0
US-10-079-754A-10 (1-58) x ABXII1394 (1-267)	
OY	1 MethysllelpehelelpeheValAPhelllemetalaleuleleualamelIeaArgAlaASP 20
Dd	44 ATGAAGATCTTATTCTTTGCTTCATTAAGGGCTCATCGACTTAGCAATATTAAAGCGTGAT 100
OY	21 SerSerelugluLysArgHisArglySAArglySLySHisHIAArgGLYTyrPhgGlnGln 40
Dd	104 TCATCTGAAGAAGAAACGTCAcAGGAAAACGAAAAAACATCAATAGAGATATTTCACA 163
OY	41 TyrGlnProTyrglnArqTYrProLeuaAnTYrProProAlaTYrProPhPro 58
Dd	164 TACCAGCATTATCAAGCATATTCACATAATTATCTCTCGGTATTCACATTCTCT 217

ID	ABX11388 standard; cDNA; 505 BP.	
XX	ABX11388;	
AC	ABX11388;	
XX		
DT	01-MAY-2003 (first entry)	
XX		
DE	Bovine histatin like polypeptide, #1, cDNA.	
XX		
KM	Cow; gene; ss; histatin; bovine; salivary secretion; oral cavity;	
KM	non-immune defence system; oral candidosis; gene therapy;	
KM	microbial infection; fungal infection; dental caries; plaque; tartar;	
KM	cystic fibrosis; systemic infection; Candida infection; mastitis;	
XX	fungicide; antibacterial.	
OS	Bos taurus.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	44..220
FT	/tag=	a
FT	/product=	"Histatin like polypeptide"
XX		
PN	US2002164625-A1.	
XX		
PD	07-NOV-2002.	
XX		
PF	19-FEB-2002; 2002US-0079754.	
XX		
PR	23-AUG-1999; 99US-150330P.	
PR	29-OCT-1999; 99US-162701P.	
PR	22-AUG-2000; 2000US-0644190.	
PR	27-OCT-2000; 2000US-0699146.	
XX		
PA	(GENE-) GENESIS RES & DEV CORP LTD.	
XX		
PI	Glenn M, Grigor MR, Molenaar AJ, Davis SR;	
XX		
DR	WPI: 2003-275306/27.	
DR	P-PSDB; ABG75791.	
XX		
PT	New histatin polynucleotides and polypeptides expressed in bovine	
PT	tissues, useful for treating bacterial or fungal infections in mammals,	
PT	e.g. infections of the oral cavity, vagina, urethra, ear or skin, or	
PT	systemic infections	
XX		
PS	Claim 1; Page 11; 17pp; English.	
XX		
CC	The invention discloses isolated histatin polynucleotides and	
CC	polypeptides, which comprise sequences expressed in bovine. Histatins are	
CC	histidine rich and specific to the salivary secretions. They are believed	
CC	to function as part of the non-immune defence system, particularly in the	
CC	oral cavity and have promise as therapeutic agents in humans with oral	
CC	candidosis. Also disclosed are compositions comprising the histatin	
CC	polypeptide or polynucleotide, and at least one component consisting of	
CC	physiological or pharmaceutical carriers or immunostimulants. The	
CC	histatin polynucleotide, polypeptide, or the cosmetic composition	
CC	comprising the polypeptide, is useful for treating (e.g. gene therapy) a	
CC	disorder in a mammal, particularly microbial or fungal infections. The	
CC	histatin polynucleotide, is also useful in genome mapping, physical	
CC	mapping or in the positional cloning of genes. Specifically, the	
CC	polypeptide or polynucleotide is useful for treating fungal or bacterial	
CC	infections of the oral cavity (e.g. dental caries, plaque or tartar),	
CC	vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis	
CC	caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating	
CC	systemic infections such as systemic Candida infection. The polypeptide	
CC	or polynucleotide is also useful for veterinary applications e.g. for	
CC	treating mastitis. The sequence presented is a cDNA encoding a bovine	
XX	histatin like polypeptide.	
XX		
SO	Sequence 505 BP; 188 A; 96 C; 67 G; 154 T; 0 other;	

Score: 316.00 Matches: 58  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 25 Gaps: 0  
US-10-079-754A-10 (1-58) x ABX11388 (1-505)  
QY 1 MetLysIlePheAlIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAsp 20  
DB 44 ATGAAGATCTTTATCTTTCTTCTTCAATTAAGCTCTCATCCATGATTAAGAGCTGAT 103  
QY 21 SerSerGluGluLysArgHisArgLysArgLysLysHisArgGlyTyrPheGlnGln 40  
DB 104 TCATCTGAAGAGAAAGTCACAGAAACGAAAAACATCATAGAGATATTTTCACAA 163  
QY 41 TTYGlnPProTyrGlnArgGlyTyrProLeuAsnTyrProProAlaTyrProPhePro 58  
DB 164 TACAGCCATATCAACGATATCCACTAAATATCTCTCGGTATCCATTCTCT 217  
RESULT 3  
ABX11391  
ID ABX11391 standard; cDNA; 604 BP.  
XX  
XX ABX11391;  
AC  
XX  
XX 01-MAY-2003 (first entry)  
DT  
XX  
XX Bovine histatin like polypeptide, #4, cDNA..  
DE  
XX  
XX Cow; gene; ss; histatin; bovine; salivary secretion; oral cavity;  
KW non-immune defence system; oral candidosis; gene therapy;  
KW microbial infection; fungal infection; dental caries; plaque; tartar;  
KW cystic fibrosis; systemic infection; Candida infection; mastitis;  
KW fungicide; antibacterial.  
XX  
XX Bos taurus.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS 154..330  
FT /\*tag= a  
FT /product= "Histatin like polypeptide"  
XX  
XX US2002164625-A1.  
XX  
XX PD 07-NOV-2002.  
XX  
XX PF 19-FEB-2002; 2002US-0079754.  
XX  
XX PR 23-AUG-1999; 99US-150330P.  
XX PR 29-OCT-1999; 99US-162701P.  
XX PR 22-AUG-2000; 2000US-0644190.  
XX PR 27-OCT-2000; 2000US-0699146.  
XX  
XX PA (GENE-) GENESIS RES & DEV CORP LTD.  
XX  
XX PI Glenn M, Grigor MR, Molenaar AJ, Davis SR,  
XX  
XX WPI; 2003-275306/27.  
XX DR P-PSDB; ABG75794.  
XX  
XX New histatin polynucleotides and polypeptides expressed in bovine  
PT tissues, useful for treating bacterial or fungal infections in mammals,  
PT e.g. infections of the oral cavity, vagina, urethra, ear or skin, or  
PT systemic infections  
XX  
XX Claim 1; Page 12; 17pp; English.  
XX  
XX PS  
XX  
XX The invention discloses isolated histatin polynucleotides and  
CC polypeptides, which comprise sequences expressed in bovine. Histatins are  
CC histidine rich and specific to the salivary secretions. They are believed  
CC to function as part of the non-immune defence system, particularly in the  
CC oral cavity and have promise as therapeutic agents in humans with oral

CC candidosis. Also disclosed are compositions comprising the histatin  
CC polypeptide or polynucleotide, and at least one component consisting of  
CC physiological or pharmaceutical carriers or immunostimulants. The  
CC histatin polynucleotide, polypeptide, or the cosmetic composition  
CC comprising the polypeptide, is useful for treating (e.g. gene therapy) a  
CC disorder in a mammal, particularly microbial or fungal infections. The  
CC histatin polynucleotide is also useful in genome mapping, physical  
CC mapping or in the positional cloning of genes. Specifically, the  
CC polypeptide or polynucleotide is useful for treating fungal or bacterial  
CC infections of the oral cavity (e.g. dental caries, plaque or tartar),  
CC vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis  
CC caused by *Pseudomonas aeruginosa*), mucosa or eye, as well as for treating  
CC systemic infections such as systemic Candida infection. The polypeptide  
CC or polynucleotide is also useful for veterinary applications e.g. for  
CC treating mastitis. The sequence presented is a cDNA encoding a bovine  
CC histatin like polypeptide.  
XX  
XX SQ Sequence 604 BP; 222 A; 112 C; 79 G; 191 T; 0 other;  
XX  
XX Alignment Scores:  
Pred. No.: 7,24e-32 Length: 604  
Score: 316.00 Matches: 58  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 25 Gaps: 0  
US-10-079-754A-10 (1-58) x ABX11391 (1-604)  
QY 1 MetLysIlePheAlIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAsp 20  
DB 154 ATGAAGATCTTTATCTTTCTTCTTCAATTAAGCTCTCATCCATGATTAAGAGCTGAT 213  
QY 21 SerSerGluGluLysArgHisArgLysArgLysLysHisArgGlyTyrPheGlnGln 40  
DB 214 TCATCTGAAGAGAAAGTCACAGAAACGAAAAACATCATAGAGATATTTTCACAA 273  
QY 41 TTYGlnPProTyrGlnArgGlyTyrProLeuAsnTyrProProAlaTyrProPhePro 58  
DB 274 TACAGCCATATCAACGATATCCACTAAATATCTCTCGGTATCCATTCTCT 327  
RESULT 4  
ABX11389  
ID ABX11389 standard; cDNA; 585 BP.  
XX  
XX ABX11389;  
AC  
XX  
XX 01-MAY-2003 (first entry)  
DT  
XX  
XX PF 19-FEB-2002; 2002US-0079754.  
XX  
XX DE Bovine histatin like polypeptide, #2, cDNA.  
XX  
XX KW Cow; gene; ss; histatin; bovine; salivary secretion; oral cavity;  
KW non-immune defence system; oral candidosis; gene therapy;  
KW microbial infection; fungal infection; dental caries; plaque; tartar;  
KW cystic fibrosis; systemic infection; Candida infection; mastitis;  
KW fungicide; antibacterial.  
XX  
XX Bos taurus.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS 149..325  
FT /\*tag= a  
FT /product= "Histatin like polypeptide"  
XX  
XX US2002164625-A1.  
XX  
XX PD 07-NOV-2002.  
XX  
XX PF 19-FEB-2002; 2002US-0079754.  
XX  
XX PR 23-AUG-1999; 99US-150330P.  
XX PR 29-OCT-1999; 99US-162701P.  
XX PR 22-AUG-2000; 2000US-0644190.





PR 27-OCT-2000; 2000US-0699146.  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
PA Glenn M, Grigor MR, Molenaar AJ, Davis SR;  
XX WPI: 2003-275306/27.  
DR P-PSDB; ABG75795.  
XX  
PT New histatin polynucleotides and polypeptides expressed in bovine  
PT tissues, useful for treating bacterial or fungal infections in mammals,  
PT e.g. infections of the oral cavity, vagina, urethra, ear or skin, or  
PT systemic infections -  
XX  
PS Claim 1; Page 12; 17pp; English.  
XX  
CC The invention discloses isolated histatin polynucleotides and  
CC polypeptides, which comprise sequences expressed in bovine. Histatins are  
CC histidine rich and specific to the salivary secretions. They are believed  
CC to function as part of the non-immune defence system, particularly in the  
CC oral cavity and have promise as therapeutic agents in humans with oral  
CC candidosis. Also disclosed are compositions comprising the histatin  
CC polypeptide or polynucleotide, and at least one component consisting of  
CC physiological or pharmaceutical carriers or immunostimulants. The  
CC histatin polynucleotide, polypeptide, or the cosmetic composition  
CC comprising the polypeptide, is useful for treating (e.g. gene therapy) a  
CC disorder in a mammal, particularly microbial or fungal infections. The  
CC histatin polynucleotide is also useful in genome mapping, physical  
CC mapping or in the positional cloning of genes. Specifically, the  
CC polypeptide or polynucleotide is useful for treating fungal or bacterial  
CC infections of the oral cavity (e.g. dental caries, plaque or tartar),  
CC vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis  
CC caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating  
CC systemic infections such as systemic Candida infection. The polypeptide  
CC or polynucleotide is also useful for veterinary applications e.g. for  
CC treating mastitis. The sequence presented is a cDNA encoding a bovine  
CC histatin like polypeptide.  
XX  
SQ Sequence 96 BP; 33 A; 25 C; 10 G; 28 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 2.45e-06 Length: 96  
Score: 112.00 Matches: 18  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 94.74% Mismatches: 0  
Query Match: 35.44% Indels: 0  
DB: 25 Gaps: 0  
US-10-079-754A-10 (1-58) x ABX11392 (1-96)  
Qy 38 PheGInGnTYrGInProTYrGLnArgTYrProleuAenTYrProProAlaTYrPro 56  
Db 1 TTTCACCAATACACGACCATATGACGATATCCACATTAATATCCCTCGCATATCA 57  
RESULT 8  
AAA48964  
ID AAA48964 standard; DNA; 552 BP.  
XX  
AC AAA48964;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human statherin DNA.  
XX  
KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;  
KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;  
KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;  
KW cancer; leukemia; adenocarcinoma; melanoma; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 73..261

PT /\*tag= a  
FT /product= Statherin  
XX  
PN MO200024779-A1.  
XX  
PD 04-MAY-2000.  
XX  
PF 22-OCT-1999; 99MO-US24046.  
XX  
PR 23-OCT-1998; 98US-0155209.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Tang YT, Corley NC, Guegler KJ, Patterson C;  
XX  
DR WPI: 2000-350699/30.  
DR P-PSDB; AAY94527.  
XX  
XX  
PT Purified polypeptide used for treating or preventing a disorder  
PT characterized by expression or activity of lysine-rich statherin  
PT proteins -  
XX  
PS Disclosure; Page 71; 75pp; English.  
XX  
CC The present invention relates to human lysine-rich statherin protein  
CC (LRSP) (AAY94526). The cDNA sequence encoding this protein was identified  
CC through analysis of a cDNA library of breast tumour tissue (BRSTN014).  
CC The LRSP sequence was found to have homology with the DNA of human  
CC statherin (the present sequence) and human basic histidine-rich protein  
CC (AAY94528). Human statherin is a phosphoprotein that acts as an  
CC inhibitor of precipitation of calcium phosphate salts in the oral cavity.  
CC The LRSP polypeptide and its antagonists may be useful for treating or  
CC preventing disorders associated with the activity of LRSP. Such  
CC disorders include autoimmune/inflammatory disorders (for example AIDS,  
CC allergies, asthma, diabetes mellitus), bacterial and fungal infection  
CC and cancers (such as leukemia, adenocarcinoma, melanoma). Antibodies to  
CC LRSP may be useful for diagnosis of the above disorders.  
XX  
SQ Sequence 552 BP; 172 A; 122 C; 78 G; 180 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 4.27e-05 Length: 552  
Score: 110.00 Matches: 27  
Percent Similarity: 61.82% Conservative: 7  
Best Local Similarity: 49.09% Mismatches: 19  
Query Match: 34.81% Indels: 2  
DB: 21 Gaps: 2  
US-10-079-754A-10 (1-58) x AAA48964 (1-552)  
Qy 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAsp 20  
Db 73 ATGAGGTTCTTGCTGCTTCATCTTGCTCTCATGGTTTCATGATTCGAGCTGAT 132  
Qy 21 SerSerGluGluIuysArgHisArgLysArgLysArgLysHisHisArgGlyTYrPheGInGln 40  
Db 133 TCATCTGAAGAGAAATTTTGGCTAGAGATTCGAGATTCGATTATGGGTAT---GGCCCT 169  
Qy 41 TYrGInProTYrGInArgTYrProleuAenTYrProProAlaTYr 55  
Db 190 TATCAGCCAGTTCACAGAACACACATTA---TACCACCAACCATATC 231  
RESULT 9  
AAV89526  
ID AAV89526 standard; cDNA; 357 BP.  
XX  
AC AAV89526;  
XX  
DT 15-FEB-1999 (first entry)  
XX  
DE EST clone CP294.  
XX  
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;





XX Human basic histidine-rich protein DNA.  
DE  
XX  
KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;  
KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;  
KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;  
KW cancer; leukemia; adenocarcinoma; melanoma; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 38..193  
FT /\*tag= a  
FT /product= Basic\_histidine-rich\_protein  
XX  
XX MO200024779-A1.  
XX  
XX 04-MAY-2000.  
XX  
XX 22-OCT-1999; 99WO-US24046.  
XX  
XX 23-OCT-1998; 98US-0155209.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX  
XX Tang YT, Corley NC, Guegler KJ, Patterson C;  
XX  
XX WPI; 2000-350699/30.  
XX  
XX P-PSDB; AAY94528.  
XX  
XX Purified polypeptide used for treating or preventing a disorder  
XX characterized by expression or activity of lysine-rich statherin  
XX proteins -  
XX  
XX  
XX Disclosure; Page 72; 75pp; English.  
XX  
XX The present invention relates to human lysine-rich statherin protein  
XX (LRSP) (AAY94526). The cDNA sequence encoding this protein was identified  
XX through analysis of a cDNA library of breast tumor tissue (BRSTN0114).  
XX The LRSP sequence was found to have homology with human statherin  
XX protein (AAY94527) and human basic histidine-rich protein (the present  
XX sequence). Human statherin is a phosphoprotein that acts as an inhibitor  
XX of precipitation of calcium phosphate salts in the oral cavity. The LRSP  
XX polypeptide and its antagonists may be useful for treating or preventing  
XX disorders associated with the activity of LRSP. Such disorders include  
XX autoimmune/inflammatory disorders (for example, AIDS, allergies, asthma,  
XX diabetes mellitus), bacterial and fungal infection and cancers (such as  
XX leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be useful  
XX for diagnosis of the above disorders.  
XX  
SQ Sequence 491 BP; 151 A; 90 C; 87 G; 163 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 0.00329 Length: 491  
Score: 95.00 Matches: 22  
Percent Similarity: 59.57% Conservative: 6  
Best Local Similarity: 46.81% Mismatches: 13  
Query Match: 30.06% Indels: 6  
DB: 21 Gaps: 2  
US-10-079-754A-10 (1-58) x AAA48965 (1-491)  
QY 1 MetyslllePhehlllePhevalPheleMetAlaleuileleuAlaMetlleArgAlaasp 20  
DB 38 ATGAAGTTTGTGTTTGTCTTAATCTTGCTCTCACTGCTTCACACTGAGCTGAT 97  
QY 21 SerSerGlulguLysArgHis---ArgLysArgLysLysHis-----His 34  
DB 98 TCACATGCAAGAGACATCATGCGTATTAAGAAATAATTCATGAAAAGCATCATTCACAT 157  
QY 35 ArgGlyTyrPheGlnGlnTyr 41  
DB 158 CGAGGCTATAGATCAATATAT 178

RESULT 12  
AAH98658  
ID AAH98658 standard; cDNA; 857 BP.  
XX  
XX AAH98658;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Human EST-derived coding sequence SEQ ID NO: 515.  
XX  
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition; ss.  
XX  
XX  
OS Homo sapiens.  
XX  
XX MO200154477-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 25-JAN-2001; 2001WO-US02687.  
XX  
XX 25-JAN-2000; 2000US-0491404.  
XX  
XX 17-JUL-2000; 2000US-0617746.  
XX  
XX 03-AUG-2000; 2000US-0631451.  
XX  
XX 15-SEP-2000; 2000US-0663870.  
XX  
XX (HYSE-) HYSQ INC.  
XX  
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
XX Cao Y, Drmanac RA, Zhang J, Werhman T;  
XX  
XX WPI; 2001-476164/51.  
XX  
XX P-PSDB; AAM23999.  
XX  
XX Isolated polypeptide for treatment of diseases, diagnostic, raising  
XX antibodies and research use -  
XX  
XX  
XX Claim 1; Page 537; 1275pp; English.  
XX  
XX The present invention provides the protein and coding sequences of novel  
XX proteins from a variety of organisms, including human, dog, cat, horse,  
XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
XX urchin and tomato. These were derived from expressed sequence tags (ESTs)  
XX from the organism of interest. They can be used in diagnostic,  
XX forensic, gene mapping, identification of mutations, to assess  
XX biodiversity and for nutritional purposes. The present sequence is a cDNA  
XX of the invention.  
XX  
SQ Sequence 857 BP; 231 A; 194 C; 169 G; 263 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 0.00676 Length: 857  
Score: 95.00 Matches: 22  
Percent Similarity: 59.57% Conservative: 6  
Best Local Similarity: 46.81% Mismatches: 13  
Query Match: 30.06% Indels: 6  
DB: 22 Gaps: 2  
US-10-079-754A-10 (1-58) x AAH98658 (1-857)  
QY 1 MetyslllePhehlllePhevalPheleMetAlaleuileleuAlaMetlleArgAlaasp 20  
DB 395 ATGAAGTTTGTGTTTGTCTTAATCTTGCTCTCACTGCTTCACACTGAGCTGAT 454  
QY 21 SerSerGlulguLysArgHis---ArgLysArgLysLysHis-----His 34  
DB 455 TCACATGCAAGAGACATCATGCGTATTAAGAAATAATTCATGAAAAGCATCATTCACAT 514  
QY 35 ArgGlyTyrPheGlnGlnTyr 41





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 13, 2003, 03:09:45 ; Search time 51 Seconds

501.965 Million cell updates/sec

Title: US-10-079-754A-10

Sequence: 1 MKIFIFVFMALILAMIRAD.....QQYQPYQRYPLNYPYPYFP 58

Scoring table: BLOSUM62

Xgapop	10.0	,	Xgapext	0.5
Ygapop	10.0	,	Ygapext	0.5
Fgapop	6.0	,	Fgapext	7.0
Delop	6.0	,	Delext	7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

### Command line parameters

```
-Q=/cgm2.1/USPTO_spool/US10079754/runat_121222003_145753_13039/app_query.fasta_1.199
-DB=Issued_Patents_NA -QFWT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
```

```
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
```

```
-USER=US10079754 @CGN_1_1_56 @runat_12122003_145753_13039 -NCPUPU=6 -ICPU=3
```

```
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGF
```

Database : Issued Patents\_NA:\*

```

1: /cgn2_6/prodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
C	1	69.5	22.0	2581	4	US-09-370-338-66	Sequence 66, Appl
C	2	66	20.9	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C	3	65	20.9	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C	4	66	20.6	2335	3	US-09-387-574-9	Sequence 9, Appl
C	5	65	20.6	2335	4	US-09-668-096-9	Sequence 9, Appl
C	6	64	20.3	4378	1	US-08-592-214A-17	Sequence 17, Appl
C	7	64	20.3	4379	3	US-09-149-976-17	Sequence 17, Appl
C	8	63.5	20.1	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C	9	63.5	20.1	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C	10	63	19.9	426	4	US-09-328-352-3819	Sequence 3819, Appl
C	11	63	19.9	750	4	US-09-359-301A-30	Sequence 30, Appl
C	12	63	19.9	857	1	US-08-308-883-1	Sequence 1, Appl

13	63	19.9	857	1	US-08-730-163-1	Sequence 1, Appl
14	63	19.9	857	3	US-08-756-799-1	Sequence 1, Appl
15	63	19.9	857	3	US-08-462-437-1	Sequence 1, Appl
16	63	19.9	858	1	US-07-962-569A-7	Sequence 7, Appl
17	62.5	19.8	1343	3	US-09-493-505-5	Sequence 5, Appl
18	62.5	19.8	1343	3	US-09-626-410-5	Sequence 5, Appl
19	62.5	19.8	1343	3	US-09-616-188-5	Sequence 5, Appl
20	62.5	19.8	1343	4	US-09-626-047-5	Sequence 5, Appl
21	62.5	19.8	1343	4	US-09-626-343-5	Sequence 5, Appl
22	62.5	19.8	1343	4	US-09-354-992-6	Sequence 6, Appl
23	62.5	19.8	1343	4	US-09-516-921-5	Sequence 6, Appl
24	62	19.6	747	4	US-09-255-991A-1185	Sequence 1185, A
25	62	19.6	1611	4	US-09-252-991A-11302	Sequence 11302, A
26	62	19.6	4100	4	US-09-620-312D-81	Sequence 81, Appl
27	61.5	19.5	2190	4	US-09-625-188-19	Sequence 19, Appl
28	61	19.3	1830121	4	US-09-557-884-1	Sequence 1, Appl
29	61	19.3	1830121	4	US-09-563-990A-1	Sequence 1, Appl
30	60.5	19.1	645	4	US-09-328-352-1899	Sequence 1899, Ap
31	60.5	19.1	1921	4	US-09-016-434-1168	Sequence 1168, Ap
32	60.5	19.1	1925	1	US-08-435-454-3	Sequence 3, Appl
33	60.5	19.1	1925	1	US-08-653-972A-3	Sequence 3, Appl
34	60.5	19.1	1925	2	US-08-913-145-5	Sequence 5, Appl
35	60.5	19.1	1925	3	US-08-870-126-3	Sequence 3, Appl
36	60.5	19.1	1925	3	US-09-344-889-5	Sequence 5, Appl
37	60.5	19.1	1925	3	US-09-444-247-3	Sequence 3, Appl
38	60.5	19.1	1925	5	PCT-US96-06231A-3	Sequence 3, Appl
39	60.5	19.1	8078	3	US-08-870-126-12	Sequence 12, Appl
40	60.5	19.1	8078	4	US-09-445-247-12	Sequence 12, Appl
41	60.5	19.1	11459	4	US-09-463-136-7	Sequence 7, Appl
42	60.5	19.1	14985	1	US-08-653-972A-6	Sequence 6, Appl
43	60.5	19.1	14985	5	PCT-US96-06231A-6	Sequence 6, Appl
44	60	19.0	708	1	US-08-078-090-24	Sequence 24, Appl
45	60	19.0	862	4	US-09-221-017B-245	Sequence 245, Appl

## ALIGNMENTS

## RESULT

; Sequence 66, Application US/09370838

; GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.

; APPLICANT: Secrist, Heather

1. TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE

CURRENT APPLICATION NUMBER: US/09/370,838

; EARLIER APPLICATION NUMBER: US

; NUMBER OF SEQ ID NOS: 289

; SEQ ID NO 66

TYPE: DNA

US-09-370-838-66

**Alignment Scores:**

**Score:**

Best Local Similarity

DB:

US-10-079-754A-10 (1-58) X US-09-370-838-66 (1-2581)

3 JLePheIJePheValPheIJeMetAlaLeuIJeLeuAlaMetIJeArgAlaAsp-serSe 22

```

Db      2236  GTCTCCGTCATCTCTTTCATTTCTCATCCCATCCGCTCGTCCGTCGTCGACCTC 2267
Oy      22    rglnglu-lyearghisarglylsrghis--hisarglytyrphelngint 41
          ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      2266  AGAGAGATCGCGGCATCTTCAGAGAGATCCCATCTTCATCTTCATCTCTTCATC 2207
Oy      41    yrglnpctyrglnargtyrproleuantyrrppropoalatyrrp 56
          :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      2206  CTCGTCCTCGTCATCTTCATCTCTTCATCTTCATCTTCGAGACTCG 2160

RESULT 2
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103, 840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Alignment Scores:
Pred. No.:      4.22e+05      Length:      4403765
Score:          66.00        Matches:      10
Percent Similarity: 58.06%    Conservative: 8
Best Local Similarity: 32.26%  Mismatches:   13
Query Match:     20.89%      Indels:       0
                       Gaps:       0

US-10-079-754A-10 (1-58) x US-09-103-840A-2 (1-4403765)
Oy      18    ArgAlaapserSerGlnGluLysArgHisArgLysArgLysHisHisArgGlyTyr 37
          ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1189176  CGTGGCATACAGACGCCACCGACGACGCGCGCGCGCGCGCGCGCGCGCGCG 1189235
Oy      38    PheGlnGlnTyrGlnProTyrGlnArgTyrPro 48
          ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1189236  CACCACCCGTCACACCTTACCGCGCGGCGCA 1189268

RESULT 3
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103, 840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA

```

```

; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.:      4.23e+05          Length:      4411529
Score:         66.00             Matches:       10
Percent Similarity:   58.06%     Conservative:    10
Best Local Similarity: 32.26%     Mismatches:    13
Query Match:        20.89%       Indels:         0
DB:                 3            Gaps:           0

US-10-079-754A-10 (1-58) x US-09-103-840A-1 (1-4411529)

QY              18 ARGALAAsPSeSerGIUGLULySArGHlEArdPLySArGLySHlEArGlyTYr 37
                |||.....:::|||||.....::|||.....::|||.....::|||
Db               1189150 CGTGGCCATACAGCAGCCCCCGCACCGGCAGGCCGCCGCCGCCGTGGCCGA 1189209

QY              38 PheIngInTyrgInProTyrgInArGYrPro 48
                |||.....::|||.....::|||.....::|||.....::|||
Db               1189210 CACCACCGCTACCACTTTACCGCGGTGGCCA 1189242


RESULT 4
US-09-387-574-9
; Sequence 9, Application US/09387574
; Patent No. 6168951
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Kinney, Tony
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Geranylgerany] Transferases
; FILE REFERENCE: BB-1239
; CURRENT APPLICATION NUMBER: US/09/387,574
; EARLIER FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: 60/098,743
; EARLIER FILING DATE: September 1, 1998
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 2335
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-387-574-9

Alignment Scores:
Pred. No.:      23.2            Length:      2335
Score:         65.00             Matches:       21
Percent Similarity:   34.15%     Conservative:    7
Best Local Similarity: 25.61%     Mismatches:    20
Query Match:        20.57%       Indels:        34
DB:                 3            Gaps:         3

US-10-079-754A-10 (1-58) x US-09-387-574-9 (1-2335)

QY              10 MetaLeuLIleAlaMeTIleArgLAAsPSeSerGIUGLUySArGHISArGLyE 29
                |||.....::|||.....::|||.....::|||.....::|||
Db               114 TTACGACTGATCCGAGCGGCGCTTTTCGCGCATATAGCGGAGAACGCGGCCGACCGGCGG 173

QY              30 Arg-----LyLVNHISHeArGLyTYrPhe----- 38
                |||.....::|||.....::|||.....::|||.....::|||
Db               174 CGGATTCCCcCGcGAGAGCTACccCACCATCGAcCcAcCTCGTTcGAGcGTgcCTCTG 233

QY              38 ----- 38
Db               234 CGGACCGGcCTCCcCGAgTCcGTTCCTCGCGcCGcCTCGGcCGCCcCGGgAAAGAcGT 293

QY              39 -----GingInTyrgInProTyrgInArGYrProLeuAnTyrgProfoALATyr-Pro 56
                .....::|||.....::|||.....::|||.....::|||
Db               294 CCTCAGCGTGACACCCAAcCTTTCTAGcGgGTCCCTCTTCCTCCCTCCCTCT---CCc 350

QY              57 PhePro 58
                |||.....::|||.....::|||.....::|||
Db               351 TTCCTT 356
```



ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/149,976  
FILING DATE: 09-SEP-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/592,214  
FILING DATE: 26-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-UD 3291  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4379 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: unsure  
LOCATION: 2095..2098  
OTHER INFORMATION: /note= "N = one or more  
OTHER INFORMATION: nucleotides."  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..4379  
OTHER INFORMATION: /note= "sequence = Arabidopsis  
OTHER INFORMATION: thaliana API gene"  
US-09-149-976-17  
Alignment Scores:  
Pred. No.: 75.8 Length: 4379  
Score: 64.00 Matches: 19  
Percent Similarity: 55.74% Conservative: 15  
Best Local Similarity: 31.15% Mismatches: 20  
Query Match: 20.25% Indels: 8  
DB: 3 Gaps: 2  
US-10-079-754A-10 (1-58) x US-09-149-976-17 (1-4379)  
QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleAlaMet----- 16  
Db 1473 ATCAATTTGTTATTTTCATTTTAAAG-TCCGAGTTTATTAATAAATCATGAGACC 1531  
QY 17 ---IleArgAlaAspSerSerGluGluLeuArgHisArgLysArgLysHisArg 35  
Db 1532 GACCTTAGTACGACATATACCAATGAGAGATCGACAGCAATCTTAAGAACCACTGT 1591  
QY 36 GlyTyrPheGlnGlnIleProTyrGlnArgTyrProLeuAsnTyrProProAlaTyr 55  
Db 1592 GGTITTCGAAACAGAGAACCGCTTACGTTTCCCTTAAGACAC-----TCTTAC 1645  
QY 56 Pro 56  
Db 1646 CCA 1648  
RESULT 8  
US-09-103-840A-2/c  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2  
Alignment Scores:  
Pred. No.: 7.31e+05 Length: 4403765  
Score: 63.50 Matches: 17  
Percent Similarity: 55.32% Conservative: 9  
Best Local Similarity: 36.17% Mismatches: 15  
Query Match: 20.09% Indels: 7  
DB: 3 Gaps: 1  
US-10-079-754A-10 (1-58) x US-09-103-840A-1 (1-4403765)  
QY 18 ArgAlaAspSerSerGluGluLeuArgHisArgLysArgLysHisArgGlyTyr 37  
Db 3771788 CGTCCGCTTCGTCGCCGCTACCGCCGACCGCGCTTCCGCGTACCGCCGACGAGAC 3771729  
QY 38 PheGlnGlnTyrGlnProTyrGlnArgTyrPro-----LeuAsnTyr 51  
Db 3771728 TACCGCGGACGCGGCGCTGCCCGCGCG-CCGCCGACCGCCATTGGACCGCGCTCA 3771670  
QY 52 ProProAlaTyrProPhePro 58  
Db 3771669 CCGCGCGGCGCGCGCTTCCCG 3771649  
RESULT 9  
US-09-103-840A-1/c  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37RV  
US-09-103-840A-1  
Alignment Scores:  
Pred. No.: 7.32e+05 Length: 4411529  
Score: 63.50 Matches: 17  
Percent Similarity: 55.32% Conservative: 9  
Best Local Similarity: 36.17% Mismatches: 15  
Query Match: 20.09% Indels: 7  
DB: 3 Gaps: 1  
US-10-079-754A-10 (1-58) x US-09-103-840A-1 (1-4411529)  
QY 18 ArgAlaAspSerSerGluGluLeuArgHisArgLysArgLysHisArgGlyTyr 37



Db 3779614 CCGGCCCGCTTCGCGCCGCGGACCGCGCGCTTCGCGCGACCGCGAGAAC 3779555  
Qy 38 PheGInGInTyRgInProTyRgInaTyRgPro-----LeuAnTyR 51  
Db 3779554 TACCGCGGACCGCGCGCTTCGCGCGCGCG-CCGCGCGGACCGCGCATTCGCGACCGCGCTCA 3779496  
Qy 52 ProProAlaTyRProProPhePro 58  
Db 3779495 CCGCGCGCGCGCGCGCTTCGCG 3779475

RESULT 10  
US-09-328-352-3819/c  
Sequence 3819, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 3819  
LENGTH: 426  
TYPE: DNA  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-3819

Alignment Scores:  
Pred. No.: 4.26 Length: 426  
Score: 63.00 Matches: 14  
Percent Similarity: 51.92% Conservative: 13  
Best Local Similarity: 26.92% Mismatches: 18  
Query Match: 19.94% Indels: 7  
Gaps: 1

US-10-079-754A-10 (1-58) x US-09-328-352-3819 (1-426)

Qy 4 PheIlePheValPheIleMeCAlaLeu-IleLeuAlaMeCAlaArgAlaSerSerGcl 23  
Db 254 TTTCTTAACCGTAATCTTGTGCTGTATTTAATTCGCTTAATGCGGTATTAATCTTA 195  
Qy 23 uGluTySaRgHsAArgLySaRg-----LysLysHsHsAArgLyTy 37  
Db 194 TTCAGTCTGCATCGCACTCCCATAGATTTATTGCTACAGTAATATCATTAAGTA 135  
Qy 37 rPheGInGInTyRgInProTyRgInaTyRPro 48  
Db 134 TTACGACGACATATCAAGTTATCAAGATAGACCG 101

RESULT 11  
US-09-359-301A-30  
Sequence 30, Application US/09359301A  
Patent No. 6426185  
GENERAL INFORMATION:  
APPLICANT: Kumagai, Monto H.  
APPLICANT: della-Cioppa, Guy R.  
APPLICANT: Erwin, Robert L.  
APPLICANT: McGee, David R.  
TITLE OF INVENTION: METHOD OF DETERMINING THE PRESENCE OF A  
TITLE OF INVENTION: TRAIT IN A PLANT BY TRANSFECTING A NUCLEIC ACID SEQUENCE OF  
TITLE OF INVENTION: A DONOR PLANT INTO A DIFFERENT HOST PLANT IN AN ANTI-SENSE  
TITLE OF INVENTION: ORIENTATION  
FILE REFERENCE: 00801013TUS04  
CURRENT APPLICATION NUMBER: US/09/359,301A  
CURRENT FILING DATE: 1999-07-21  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 30  
LENGTH: 750  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana

US-09-359-301A-30  
Alignment Scores:  
Pred. No.: 9.25 Length: 750  
Score: 63.00 Matches: 13  
Percent Similarity: 58.33% Conservative: 1  
Best Local Similarity: 54.17% Mismatches: 10  
Query Match: 19.94% Indels: 0  
Gaps: 0

US-10-079-754A-10 (1-58) x US-09-359-301A-30 (1-750)

Qy 35 ArgGlyTyRPhGInGInTyRgInProTyRgInaTyRProLeuAAnTyRProAla 54  
Db 199 CAGGATATCTCTCAACAAGTTATCTTCTCAGCAGAGATATCTTCCACCGTACGCGCTCA 258  
Qy 55 TyRProPhePro 58  
Db 259 TATCTCCACCA 270

RESULT 12  
US-08-308-883-1  
Sequence 1, Application US/08308883  
Patent No. 5576300  
GENERAL INFORMATION:  
APPLICANT: Mukerji, P.  
APPLICANT: Prieto, P. A.  
APPLICANT: Seo, A. E.-Y.  
APPLICANT: Baxter, J. H.  
APPLICANT: Cummings, R. D.  
TITLE OF INVENTION: Method for Inhibition of Human Rotavirus Infection.  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lonnie R. Drayer  
ADDRESSEE: ROSS Products Division  
ADDRESSEE: Abbott Laboratories  
STREET: 625 Cleveland Avenue  
CITY: Columbus  
STATE: Ohio  
COUNTRY: United States  
ZIP: 43215  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh System 7.1  
SOFTWARE: ClarisWorks 1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/308,883  
FILING DATE: 16-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (614) 624-3774  
TELEFAX: (614) 624-3074  
TELEX: No. 5576300e  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 857 base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: Human milk kappa-casein  
HYPOTHETICAL: No  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE: Human  
ORGANISM: Homo sapiens  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE: Adult  
HAPLOTYPE:  
TISSUE TYPE: Mammary gland

CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE: Human Mammary Gland  
LIBRARY:  
CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 45...593  
IDENTIFICATION METHOD: DNA sequencing and restriction analysis  
OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the human  
PUBLICATION INFORMATION:  
AUTHORS: L. Hansson et al  
TITLE: DNA Encoding Kappa-Casein, Process for Obtaining the Protein and Use The  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER: PCT/WO93/15196  
FILING DATE: 25-JAN-1993  
PUBLICATION DATE: 05-AUG-1993  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-308-863-1

Alignment Scores:  
Pred. No.: 11.1 Length: 857  
Score: 63.00 Matches: 17  
Percent Similarity: 50.85% Conservative: 13  
Best Local Similarity: 28.81% Mismatches: 25  
Query Match: 19.94% Indels: 4  
DB: 1 Gaps: 3

US-10-079-754A-10 (1-58) x US-08-308-863-1 (1-857)

Qy 1 MetLysIlePheIlePheValPhe---IleMetAlaIleuIleuAlaMetIleArgAla 19  
Db 45 ATGAAGAGTTTCTTCTAGTGTGCAATGCCCTGSCATTACCCCTTTTGGCTGTG 104

Qy 20 AspSerSerGluGluLysArgHisArgLysLysHisArgGlyTyrPheGln 39  
Db 105 GAGTTTCAAAACCAACCAACCAACGATGCATGATGATGATGAAGACCTTTCTATCAG 164

Qy 40 GlnTyrGlnProTyrGlnArgTyrProLeuAsnTyr---ProProAlaTyrProPhe 57  
Db 165 AAACAGCTCCATAT-----GTCCCAATGTATTATGTGCCAATAGCTATCCTTAT 215

RESULT 13  
US-08-730-163-1  
Sequence 1, Application US/08730163  
Patent No. 5712250  
GENERAL INFORMATION:  
APPLICANT: Mukerji, P.  
APPLICANT: Prieto, P. A.  
APPLICANT: Seo, A. E.-Y.  
APPLICANT: Baxter, J. H.  
APPLICANT: Cummings, R. D.  
TITLE OF INVENTION: Product for inhibition of Human Rotavirus Infection.  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lonnie R. Drayer  
ADDRESSEE: ROSS Products Division  
STREET: 625 Cleveland Avenue  
CITY: Columbus  
STATE: Ohio  
COUNTRY: United States  
ZIP: 43215  
COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb storage (B) COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh System 7.1(D) SOFTWARE: ClarisWorks 1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/730,163  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/308,882  
FILING DATE: 16-SEP-1994  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (614) 624-3774  
TELEFAX: (614) 624-3074  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 857 base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: Human milk kappa-casein  
HYPOTHETICAL: No  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE: Human  
ORGANISM: Homo sapiens  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE: Adult  
HAPLOTYPE:  
TISSUE TYPE: Mammary gland  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE: Human Mammary Gland  
LIBRARY:  
CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 45...593  
IDENTIFICATION METHOD: DNA sequencing and restriction analysis  
OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the human  
PUBLICATION INFORMATION:  
AUTHORS: L. Hansson et al  
TITLE: DNA Encoding Kappa-Casein, Process for Obtaining the Protein and Use The  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER: PCT/WO93/15196  
FILING DATE: 25-JAN-1993  
PUBLICATION DATE: 05-AUG-1993  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-730-163-1

Alignment Scores:  
Pred. No.: 11.1 Length: 857  
Score: 63.00 Matches: 17  
Percent Similarity: 50.85% Conservative: 13  
Best Local Similarity: 28.81% Mismatches: 25  
Query Match: 19.94% Indels: 4  
DB: 1 Gaps: 3

US-10-079-754A-10 (1-58) x US-08-730-163-1 (1-857)

Qy 1 MetLysIlePheIlePheValPhe---IleMetAlaIleuIleuAlaMetIleArgAla 19  
Db 45 ATGAAGAGTTTCTTCTAGTGTGCAATGCCCTGSCATTACCCCTTTTGGCTGTG 104

QY 20 AepSerSerGluGluysArgHisArgLysArgLysLysHisArgGlyTyrPheGln 39  
:::|||||:::|||||  
Db 105 GAGGTTCAAAACAGAAACACAGCATGCATGAGATGAAAGACATTCATATCAG 164  
QY 40 GlnTyrGlnProTyrGlnArgTyrProLeuAsnTyr---ProProAlaTyrProPhe 57  
:::|||||:::|||||  
Db 165 AAAACAGCTCCATAT-----GTCCCAATGATATTATGTCCAAAATAGCTATCTTAT 215

RESULT 14  
US-08-256-799-1  
; Sequence 1, Application US/08256799  
; Patent No. 6222094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMOVIST, Mats  
; APPLICANT: BERGSTROM, Sven  
; APPLICANT: HERNNELL, Olle  
; APPLICANT: TOERNELL, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS FOR  
; TITLE OF INVENTION: OBTAINING THE PROTEIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROMDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 06-DEC-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28, 005  
; REFERENCE/DOCKET NUMBER: HANSSON=1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 857 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 45..104  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 13..44  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: 594..848

US-08-256-799-1  
Alignment Scores:  
Pred. No.: 11.1 Length: 857  
Score: 63.00 Matches: 17  
Percent Similarity: 50.85% Conservative: 13  
Best Local Similarity: 28.81% Mismatches: 25  
Query Match: 19.94% Indels: 4  
DB: 3 Gaps: 3

US-10-079-754A-10 (1-58) x US-08-256-799-1 (1-857)  
QY 1 MetLysIlePheIlePheValPhe---IleMetAlaLeuIleLeuAlaMetIleArgAla 19  
|||||:::|||||  
Db 45 ATGAAGAGTTTCTCTAGTTGTCAATGCCTGGCATTAACCCCTTTTGGCTGTG 104  
QY 20 AepSerSerGluGluysArgHisArgLysArgLysLysHisArgGlyTyrPheGln 39  
:::|||||:::|||||  
Db 105 GAGGTTCAAAACAGAAACACAGCATGCATGAGATGAAAGACATTCATATCAG 164  
QY 40 GlnTyrGlnProTyrGlnArgTyrProLeuAsnTyr---ProProAlaTyrProPhe 57  
:::|||||:::|||||  
Db 165 AAAACAGCTCCATAT-----GTCCCAATGATATTATGTCCAAAATAGCTATCTTAT 215

RESULT 15  
US-08-462-437-1  
; Sequence 1, Application US/08462437  
; Patent No. 6232094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMOVIST, Mats  
; APPLICANT: BERGSTROM, Sven  
; APPLICANT: HERNNELL, Olle  
; APPLICANT: TOERNELL, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS  
; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROMDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,437  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28, 005  
; REFERENCE/DOCKET NUMBER: HANSSON=1A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELETYPE: 248633  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 857 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens

```
FEATURE:
NAME/KEY: CDS
LOCATION: 45..593
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 45..593
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 45..104
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 13..44
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 594..848
US-08-462-437-1
```

```
Alignment Scores:
Pred. No.: 11.1 Length: 857
Score: 63.00 Matches: 17
Percent Similarity: 50.85% Conservative: 13
Best Local Similarity: 28.81% Mismatches: 25
Query Match: 19.94% Indels: 4
DB: 3 Gaps: 3
```

US-10-079-754A-10 (1-58) x US-08-462-437-1 (1-857)

```
QY 1 MetLysIlePheIlePheValPhe---IleMetAlaLeuIleLeuAlaMetIleArgAla 19
   |||||
Db 45 ATGAAGAGTTTCTTAGTGTCAATGCCCTGGCATTACCCCTTTTGGCTGTG 104
   |||||

QY 20 AspSerSerGluGluLysArgHisArgLysArgLysLysHisLysArgLysPheGln 39
   ::|||
Db 105 GAGGTTCAAAACCAAAACCAACCAACGATGCATGAGATGATGAAGAACCATTCATCAG 164
   ::|||

QY 40 GlnTyrGlnProTyrGlnArgTyrProLeuAsnTyr---ProProAlaTyrProPhe 57
   ::|||
Db 165 AAACACAGCTCCATAT-----GTCCCAATGTATTATGTGCCAATAGCTATCCTTAT 215
   ::|||
```

Search completed: December 13, 2003, 04:44:10  
Job time : 1240 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 13, 2003, 03:14:26 ; Search time 238 Seconds  
(without alignments)  
809.953 Million cell updates/sec

Title: US-10-079-754A-10  
Perfect score: 316  
Sequence: 1 MKRIFRIFMALILMIRAD.....QQYQYRYPPLNYPAYFPF 58

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2201672 seqs, 166179599 residues  
Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n-model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US10079754/runat\_12122003\_145753\_13072/app\_query.fasta\_1.199  
-DB=PubMed Applications\_NA -QWt=faetbap -SUFFIX=rmpb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10079754 @CGN 1.1.352 @runat\_12122003\_145753\_13072  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEJOINT -NEG\_SCORES=0 -WAIT -DSBIOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-GAPOP=6 -GAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US09A\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	316	100.0	267	14	US-10-079-754A-15	Sequence 15, Appl
2	316	100.0	267	14	US-10-079-623-200	Sequence 200, App
3	316	100.0	505	14	US-10-079-754A-1	Sequence 1, Appl
4	316	100.0	604	14	US-10-079-754A-4	Sequence 4, Appl
5	315	99.7	585	14	US-10-079-754A-2	Sequence 2, Appl
6	277	87.7	525	14	US-10-079-754A-6	Sequence 6, Appl
7	170.5	54.0	869	14	US-10-079-754A-3	Sequence 3, Appl
8	112	35.4	96	14	US-10-079-754A-5	Sequence 5, Appl
9	94.5	29.9	438	11	US-09-992-600A-5	Sequence 5, Appl
10	94.5	29.9	438	11	US-09-924-340-5	Sequence 5, Appl
11	94.5	29.9	438	13	US-09-992-095B-5	Sequence 5, Appl
12	94.5	29.9	438	13	US-10-154-678-5	Sequence 5, Appl
13	94.5	29.9	438	13	US-09-999-570-5	Sequence 5, Appl
14	94.5	29.9	438	13	US-10-001-142-5	Sequence 5, Appl
15	94.5	29.9	438	15	US-10-000-489-5	Sequence 5, Appl
16	94.5	29.9	438	15	US-10-000-986-5	Sequence 5, Appl
17	78.5	24.8	2223	15	US-10-128-714-1032	Sequence 5, Appl
18	78.5	24.8	2223	15	US-10-128-714-2032	Sequence 1032, Ap
19	78.5	24.8	2223	15	US-10-128-714-6032	Sequence 2032, Ap
20	78.5	24.8	2223	15	US-10-128-714-7032	Sequence 7032, Ap
21	78.5	24.8	4223	15	US-10-128-714-32	Sequence 32, Appl
22	78.5	24.8	4223	15	US-10-128-714-5032	Sequence 5032, Ap
23	69.5	22.0	449	11	US-09-918-995-20169	Sequence 20169, A
24	69.5	22.0	517	10	US-09-920-300A-313	Sequence 313, App
25	69.5	22.0	517	13	US-10-099-926-313	Sequence 313, App
26	69.5	22.0	517	14	US-10-033-528-113	Sequence 31, App
27	69.5	22.0	2581	10	US-09-738-973-66	Sequence 66, Appl
28	69.5	22.0	2581	10	US-09-854-133-66	Sequence 66, Appl
29	69.5	22.0	2581	15	US-10-144-649A-66	Sequence 66, Appl
30	67.5	21.4	3673778	13	US-10-312-841-2	Sequence 2, Appl
31	67	21.2	612	13	US-10-027-632-116142	Sequence 116142,
32	67	21.2	612	14	US-10-027-632-116142	Sequence 116142,
33	66.5	21.0	5310	15	US-10-172-086-11	Sequence 11, Appl
34	65.5	20.7	5239	13	US-10-311-455-245	Sequence 245, Appl
35	65.5	20.7	3673778	13	US-10-312-841-1	Sequence 1, Appl
36	64.5	20.4	1289	10	US-09-938-482A-4846	Sequence 4846, Ap
37	64.5	20.4	11910	15	US-10-156-761-2879	Sequence 2879, Ap
38	64.5	20.4	125746	15	US-10-156-761-15102	Sequence 15102, A
39	64.5	20.4	9025608	15	US-10-156-761-1	Sequence 1, Appl
40	64	20.3	6385	13	US-10-240-453-60	Sequence 60, Appl
41	64	20.3	6385	15	US-10-239-676-58	Sequence 58, Appl
42	63.5	20.1	8413	13	US-10-240-485-59	Sequence 49, Appl
43	63.5	20.1	8456	9	US-10-311-455-1949	Sequence 1949, Ap
44	63	19.9	483	9	US-09-864-761-11010	Sequence 11010, A
45	63	19.9	585	15	US-10-142-077-32	Sequence 32, Appl

## ALIGNMENTS

RESULT 1  
US-10-079-754A-15  
; Sequence 15, Application US/10079754A  
; Publication No. US20020164625A1  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Grigor, Murray R.  
; APPLICANT: Molenaar, Adrian J.  
; APPLICANT: Davis, Stephen R.  
; TITLE OF INVENTION: Compositions Isolated from Bovine  
; FILE REFERENCE: 11000.1068  
; CURRENT APPLICATION NUMBER: US/10/079, 754A  
; PRIOR FILING DATE: 2002-02-19  
; PRIOR APPLICATION NUMBER: US 09/699,146  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US 60,162,701  
; PRIOR FILING DATE: 1999-10-29  
; PRIOR APPLICATION NUMBER: US 09/644,190  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: US 60,150,330  
; PRIOR FILING DATE: 1999-08-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 15
; LENGTH: 267
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-754A-15

Alignment Scores:
Pred. No.:      8.94e-37      Length:      267
Score:          316.00       Matches:      58
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    100.00%      Indels:        0
DB:             14          Gaps:           0

US-10-079-754A-10 (1-58) x US-10-079-754A-15 (1-267)

QY      1 MetlysiIlePheIlePheValPheIleMeAlaleuIleuAlaMeIleArgAlaasp 20
Db      44 ATGAAGATCTTATCTTGTCTTCATTATGCTCTCAATCCATGCGCATATAGAGCTGAT 103

QY      21 SerSerGluGluYsaArgHisArgLySarGlyshHisArgGlyTYrPheGlnGln 40
Db      104 TCATCTGAAGAAACGTCACAGAAACGAAACATCATAGAGATATTTTCAACAA 163

QY      41 TyrlInProTYrGlnArgTYrProleuAsnTYrProAlaTYrProPhePro 58
Db      164 TACCAGCATATCAACGATATCCATAATTATCTCTCGGTATCCATTTCT 217

RESULT 2
US-10-079-623-200
; Sequence 200, Application US/10079623
; Publication No. US20020169302A1
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molesaar, Adrian J.
; TITLE OF INVENTION: Compositions isolated from bovine
; TITLE OF INVENTION: Mammary gland and methods for their use.
; FILE REFERENCE: 11000.10443
; CURRENT APPLICATION NUMBER: US/10/079.623
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 200
; LENGTH: 267
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-623-200

Alignment Scores:
Pred. No.:      8.94e-37      Length:      267
Score:          316.00       Matches:      58
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    100.00%      Indels:        0
DB:             14          Gaps:           0

US-10-079-754A-10 (1-58) x US-10-079-623-200 (1-267)

QY      1 MetlysiIlePheIlePheValPheIleMeAlaleuIleuAlaMeIleArgAlaasp 20
Db      44 ATGAAGATCTTATCTTGTCTTCATTATGCTCTCAATCCATGCGCATATAGAGCTGAT 103

QY      21 SerSerGluGluYsaArgHisArgLySarGlyshHisArgGlyTYrPheGlnGln 40
Db      104 TCATCTGAAGAAACGTCACAGAAACGAAACATCATAGAGATATTTTCAACAA 163

QY      41 TyrlInProTYrGlnArgTYrProleuAsnTYrProAlaTYrProPhePro 58
Db      164 TACCAGCATATCAACGATATCCATAATTATCTCTCGGTATCCATTTCT 217

RESULT 3
5
```

```
US-10-079-754A-1
; Sequence 1, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molesaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions isolated from bovine
; TITLE OF INVENTION: Mammary gland and methods for their use
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079.754A
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-754A-1

Alignment Scores:
Pred. No.:      2.13e-36      Length:      505
Score:          316.00       Matches:      58
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    100.00%      Indels:        0
DB:             14          Gaps:           0

US-10-079-754A-10 (1-58) x US-10-079-754A-1 (1-505)

QY      1 MetlysiIlePheIlePheValPheIleMeAlaleuIleuAlaMeIleArgAlaasp 20
Db      44 ATGAAGATCTTATCTTGTCTTCATTATGCTCTCAATCCATGCGCATATAGAGCTGAT 103

QY      21 SerSerGluGluYsaArgHisArgLySarGlyshHisArgGlyTYrPheGlnGln 40
Db      104 TCATCTGAAGAAACGTCACAGAAACGAAACATCATAGAGATATTTTCAACAA 163

QY      41 TyrlInProTYrGlnArgTYrProleuAsnTYrProAlaTYrProPhePro 58
Db      164 TACCAGCATATCAACGATATCCATAATTATCTCTCGGTATCCATTTCT 217

RESULT 4
US-10-079-754A-4
; Sequence 4, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molesaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions isolated from bovine
; TITLE OF INVENTION: Mammary gland and methods for their use
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079.754A
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 146
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-754A-4

Alignment Scores:
Pred. No.:      2.13e-36      Length:      146
Score:          316.00       Matches:      58
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    100.00%      Indels:        0
DB:             14          Gaps:           0

US-10-079-754A-10 (1-58) x US-10-079-754A-4 (1-146)

QY      1 MetlysiIlePheIlePheValPheIleMeAlaleuIleuAlaMeIleArgAlaasp 20
Db      44 ATGAAGATCTTATCTTGTCTTCATTATGCTCTCAATCCATGCGCATATAGAGCTGAT 103

QY      21 SerSerGluGluYsaArgHisArgLySarGlyshHisArgGlyTYrPheGlnGln 40
Db      104 TCATCTGAAGAAACGTCACAGAAACGAAACATCATAGAGATATTTTCAACAA 163

QY      41 TyrlInProTYrGlnArgTYrProleuAsnTYrProAlaTYrProPhePro 58
Db      164 TACCAGCATATCAACGATATCCATAATTATCTCTCGGTATCCATTTCT 217

RESULT 5
5
```

```
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-754A-4

Alignment Scores:
Pred. No.: 2,71e-36      Length: 604
Score: 316.00           Matches: 58
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 14                  Gaps: 0

US-10-079-754A-10 (1-58) x US-10-079-754A-4 (1-604)

QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAsp 20
DB 154 ATGAAGATCTTATCTTTCTTCATTATGGCTCTCATCGCCATGATTAGAGCTGAT 213
QY 21 SerSerGIuLulYeaRHisArgLysArgLysLysHisHisArgGlyTyrPheGlnGln 40
DB 214 TCATCTGAAGAAAGACGTACAGAGAAACGAAACATCATATAGAGATATTTTCAACA 273
QY 41 TyrGlnProTyrGlnArgTyrProLeuAsnTyrProProAlaTyrProPhePro 58
DB 274 TACCGACCATATCAACGATATCCACTAATTTATCTCTCGATTCATTCTCT 327

RESULT 5
US-10-079-754A-2
; Sequence 2, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions Isolated from Bovine
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079,754A
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/699,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60,162,701
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/644,190
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 60,150,330
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-754A-2

Alignment Scores:
Pred. No.: 3.63e-36      Length: 585
Score: 315.00           Matches: 57
Percent Similarity: 100.00%  Conservative: 1
Best Local Similarity: 98.28%  Mismatches: 0
Query Match: 99.68%      Indels: 0
DB: 14                  Gaps: 0

US-10-079-754A-10 (1-58) x US-10-079-754A-2 (1-585)

QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAsp 20
DB 149 ATGAAGATCTTATCTTTCTTCATTATGGCTCTCATCGCCATGATTAGAGCTGAT 208
```

```
QY 21 SerSerGIuLulYeaRHisArgLysArgLysLysHisHisArgGlyTyrPheGlnGln 40
DB 209 TCATCTGAAGAAAGACGTACAGAGAAACGAAACATCATATAGAGATATTTTCAACA 268
QY 41 TyrGlnProTyrGlnArgTyrProLeuAsnTyrProProAlaTyrProPhePro 58
DB 269 TACCGACCATATCAACGATATCCACTAATTTATCTCTCGATTCATTCTCT 322

RESULT 6
US-10-079-754A-6
; Sequence 6, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions Isolated from Bovine
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079,754A
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/699,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60,162,701
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/644,190
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 60,150,330
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-754A-6

Alignment Scores:
Pred. No.: 1.01e-30      Length: 525
Score: 277.00           Matches: 57
Percent Similarity: 82.61%  Conservative: 0
Best Local Similarity: 82.61%  Mismatches: 1
Query Match: 87.66%      Indels: 12
DB: 14                  Gaps: 1

US-10-079-754A-10 (1-58) x US-10-079-754A-6 (1-525)

QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAsp 20
DB 43 ATGAAGATCTTATCTTTCTTCATTATGGCTCTCATCGCCATGATTAGAGCTGAT 102
QY 21 SerSerGIuLulYeaRHisArgLysArgLysLysHisHisArgGlyTyrPheGlnGln 35
DB 103 TCATCTGAAGAAAGACGTACAGAGAAACGAAACATCATATGATAGCTCCACA 161
QY 36 -----GlyTyrPheGlnGlnTyrGlnProTyrProLeu 49
DB 162 ATTCTTACTAATACAAAGAGATATTTTCAACAATATCCAGCATATCAAGATATCACTA 221
QY 50 AsnTyrProProAlaTyrProPhePro 58
DB 222 AATTATCTCTCTGCGATTCATTCTCT 248

RESULT 7
US-10-079-754A-3
; Sequence 3, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
```





US-10-079-754A-10 (1-58) X US-09-992-600A-5 (1-438)

```

OY  _ MelYstIlePheIlePValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaasp 20
Db  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 ATGAAGTTTTTGTCTTGCTTAGCTGGCTTCATATTCATATTCATATTAAGCGCTGAT 143
OY  21 SerSerGIuGIuIbysArgHIsArgIySArgIySLySHIsHIsArgGlyTyrPheGlnGln 40
Db  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
144 TCACATGAAAAGACATCATGGGCTTTAGAGAGAAATTCGAT-----185
OY  41 TyrGlnProTyrGlnArgTyrProLeuAsnTyrProProAlaTyr 55
Db  :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
186 ----GAAAGCATCATTCATCCATATTCACACATCACTACTCTTTT 227

```

RESULT 10  
US-09-924

```

Sequence 5, Application US/09924340
Publication No. US20030027248A1
GENERAL INFORMATION:
APPLICANT: Bejani, Stephane
APPLICANT: Tanaka, Hiroaki
FILE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US2.REG
CURRENT APPLICATION NUMBER: US/09/924,340
PRIORITY FILING DATE: 2001-08-06
PRIORITY APPLICATION NUMBER: US 60/305,456
PRIORITY FILING DATE: 2001-07-13
PRIORITY APPLICATION NUMBER: US 60/302,277
PRIORITY FILING DATE: 2001-06-29
PRIORITY APPLICATION NUMBER: US 60/298,698
PRIORITY FILING DATE: 2001-06-15
PRIORITY APPLICATION NUMBER: US 60/293,574
PRIORITY FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: UPatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
NAME/KEY: CDS
LOCATION: 84..317
NAME/KEY: 3'UTR
LOCATION: 318..438
NAME/KEY: polyA_signal
LOCATION: 397..402
NAME/KEY: polyA_site
LOCATION: 423..438
US-09-924-340-5

```

**Alignment Scores:**

Pred. No.:	0.000224	438
Score:	94.50	18
Percent Similarity:	56.36%	Conservative: 13
Best Local Similarity:	32.73%	Mismatches: 17
Query Match:	29.91%	Indels: 7
DB:	11	Gaps: 1

US-10-079-754A-10 (1-58) X US-09-924-340-5 (1-438)

[illegible]

RESULT 11

```

US-03-992-095B-5
: Sequence 5, Application US/09992095B
: Publication No. US20030157485A1
: GENERAL INFORMATION:
: APPLICANT: Benjamin, Stephanie
: APPLICANT: Tanaka, Hiroaki
: TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
: FILE REFERENCE: 91. US. DIV
: CURRENT APPLICATION NUMBER: US/09/992, 095B
: PRIORITY FILING DATE: 2003-02-20
: PRIOR APPLICATION NUMBER: US 09/924, 340
: PRIOR FILING DATE: 2001-08-06
: PRIOR APPLICATION NUMBER: PCT/IB01/01715
: PRIOR FILING DATE: 2001-08-06
: PRIOR APPLICATION NUMBER: US. 60/305, 456
: PRIOR FILING DATE: 2001-07-13
: PRIOR APPLICATION NUMBER: US 60/302, 277
: PRIOR FILING DATE: 2001-06-29
: PRIOR APPLICATION NUMBER: US 60/298, 698
: PRIOR FILING DATE: 2001-06-15
: PRIOR APPLICATION NUMBER: US 60/293, 574
: PRIOR FILING DATE: 2001-05-25
: NUMBER OF SEQ ID NOS: 112
: SOFTWARE: Jpatent
: SEQ ID NO 5

```

**Alignment Scores:**

Pred. No.:	0.000224	Length:	438
Score:	94.50	Matches:	18
Percent Similarity:	56.36%	Conservative:	13
Best Local Similarity:	32.73%	Mismatches:	17
Query Match:	29.91%	Indels:	7
DB:	13	Gaps:	1

US-10-079-754A-10 (1-58) x US-09-992-095B-5 (1-438)

```

QY      1 MetysIILepheleValPheilemetAlaleuileuAlametiLeArgAsp 20
        ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      84 ATAAAGTTTTTGTCTTGTCCTTAAGCTTTGGCTTCATGATTTCAGATTAGCGCTGAT 145
QY      21 SerSerGIUGLylsArghisArgylsArgylsVshisArGslTyTPheGlInl 40
        ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      144 TCACATGAAGAAGACATCATCGGATATGAAAGAAATTCAT----- 185
QY      41 TyrglnProTyrglnArgTyProleuanstTyProproAlatyr 55
        ::::: :||| :||| :||| :||| :||| :||| :|||
Db      186 ---GAAAAGCATCATTCATTCATTCACACTACTACCACTTTT 227

RESULT 12
US-10-154-678-5
; Sequence 5, Application US/10154678
; Publication No. US20030162186A1
```



```
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: Jpatent
; SEQ ID NO 5
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..83
; NAME/KEY: CDS
; LOCATION: 84..317
; NAME/KEY: 3'UTR
; LOCATION: 318..438
; NAME/KEY: polyA_signal
; LOCATION: 397..402
; NAME/KEY: polyA_site
; LOCATION: 423..438
; US-10-001-142-5

Alignment Scores:
Pred. No.: 0.000224 Length: 438
Score: 94.50 Matches: 18
Percent Similarity: 56.36% Conservative: 13
Best Local Similarity: 32.73% Mismatches: 17
Query Match: 29.91% Indels: 7
DB: 13 Gaps: 1

US-10-079-754A-10 (1-58) x US-10-001-142-5 (1-438)
QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleAlaMetIleArgAlaAsp 20
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 84 ATGAAGTTTGTGCTTTCCTTAGCTTGGCTCTCATGATTTCCATGATTTAGCGCTGAT 143
QY 21 SerSerGluGluLysArgHisArgLysArgLysLysHisArgGlyTyrPheGlnGln 40
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 144 TCACATGAAAGAGACATCATCGGTATGAGAAATTCAT----- 185
QY 41 TyrGlnProTyrGlnArgTyrProLeuAsnTyrProProAlaTyr 55
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 186 ---GAAAGACATCATTCATCATATACACATCACTACACACTTTT 227

RESULT 15
US-10-000-489-5
; Sequence 5, Application US/10000489
; Publication No. US20030092011A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephanie
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91 US6 DIV
; CURRENT APPLICATION NUMBER: US/10/000,489
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: Jpatent
; SEQ ID NO 5
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
```

```
; LOCATION: 1..83
; NAME/KEY: CDS
; LOCATION: 84..317
; NAME/KEY: 3'UTR
; LOCATION: 318..438
; NAME/KEY: polyA_signal
; LOCATION: 397..402
; NAME/KEY: polyA_site
; LOCATION: 423..438
; US-10-000-489-5

Alignment Scores:
Pred. No.: 0.000224 Length: 438
Score: 94.50 Matches: 18
Percent Similarity: 56.36% Conservative: 13
Best Local Similarity: 32.73% Mismatches: 17
Query Match: 29.91% Indels: 7
DB: 15 Gaps: 1

US-10-079-754A-10 (1-58) x US-10-000-489-5 (1-438)
QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleAlaMetIleArgAlaAsp 20
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 84 ATGAAGTTTGTGCTTTCCTTAGCTTGGCTCTCATGATTTCCATGATTTAGCGCTGAT 143
QY 21 SerSerGluGluLysArgHisArgLysArgLysLysHisArgGlyTyrPheGlnGln 40
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 144 TCACATGAAAGAGACATCATCGGTATGAGAAATTCAT----- 185
QY 41 TyrGlnProTyrGlnArgTyrProLeuAsnTyrProProAlaTyr 55
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 186 ---GAAAGACATCATTCATCATATACACATCACTACACACTTTT 227

Search completed: December 13, 2003, 04:28:38
Job time : 249 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 13, 2003, 03:01:44 ; Search time 24 Seconds

(without alignments)  
232.408 Million cell updates/sec

Title: US-10-079-754A-10

Perfect score: 316

Sequence: 1 MKIFIFVFMALILMIRAD.....QQYQYQRYPLNPPYPPFP 58

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	34.8	62	1	SBHUP
2	102	32.3	57	1	A32541
3	95	30.1	51	1	B32541
4	81.5	25.8	61	1	SBMOP1
5	75.5	23.9	181	2	UC5233
6	75.5	23.9	182	2	A44157
7	72	22.8	1271	2	T24008
8	70.5	22.3	42	2	UK0224
9	66.5	21.0	180	2	UC4955
10	66	20.9	310	2	T06211
11	66	20.9	386	2	T39588
12	66	20.9	733	2	TS1434
13	65.5	20.7	447	2	F82862
14	65.5	20.7	1203	2	T21275
15	65	20.6	260	2	S18350
16	65	20.6	282	2	T26635
17	65	20.6	347	2	T05737
18	64.5	20.4	533	2	T39025
19	64	20.3	380	2	E88421
20	64	20.3	400	2	T24258
21	63.5	20.1	177	2	AG1831
22	63.5	20.1	293	2	S07365
23	63	19.9	98	2	T02437
24	63	19.9	182	2	KKHU
25	63	19.9	684	2	T01267
26	62	19.6	670	2	S47042
27	60.5	19.1	215	2	S46966
28	60.5	19.1	226	1	KBHU
29	60.5	19.1	504	2	A57215

## ALIGNMENTS

30	60	19.0	269	2	T42919	protein tio - atel
31	60	19.0	662	2	T23757	hypothetical prote
32	60	19.0	1074	2	T24877	hypothetical prote
33	60	19.0	1076	2	T24887	hypothetical prote
34	59.5	18.8	1464	2	S58984	development protei
35	59	18.7	409	2	A43256	mobiliation prote
36	59	18.7	473	2	A54691	octamer-binding pr
37	59	18.7	894	2	F84870	hypothetical prote
38	58.5	18.5	413	2	T04520	hypothetical prote
39	58.5	18.5	1870	2	S37671	MHC class III hist
40	58.5	18.5	1872	2	S36152	MHC class III hist
41	58.5	18.5	2139	2	S46404	vitellogenin - yel
42	58.5	18.5	2142	2	B35098	MHC class III hist
43	58	18.4	113	2	S66936	probable membrane
44	58	18.4	156	2	S77873	probable phosphat
45	58	18.4	433	2	B47041	hypothetical prote

## RESULT 1

SBHUP  
statherin precursor [validated] - human  
C:Species: Homo sapiens (man)  
C:Date: 24-Apr-1984 #sequence revision 30-Jun-1993 #text\_change 08-Dec-2000  
C:Accession: J01015; A27308; B27489; A03288; A32524  
R:Sabbatini, L.M.; He, Y.Z.; Azen, E.A.  
Gene 89, 245-251, 1990  
A>Title: Structure and sequence determination of the gene encoding human salivary statherin  
A:Reference number: J01015; MUID:90323623; PMID:2373369  
A:Accession: J01015  
A:Molecule type: DNA  
A:Residues: 1-62 <SA2>  
A:Cross-references: GB:M31077  
R:Sabbatini, L.M.; Carlock, L.R.; Johnson, G.W.; Azen, E.A.  
Am. J. Hum. Genet. 41, 1048-1060, 1987  
A>Title: cDNA cloning and chromosomal localization (4q11-13) of a gene for statherin, a  
A:Reference number: A27308; MUID:88074310; PMID:3502720  
A:Accession: A27308  
A:Molecule type: mRNA  
A:Residues: 1-62 <SAB>  
A:Cross-references: GB:M32639; NID:G338504; PIDN:AAA60593.1; PID:G338506  
R:Dickinson, D.P.; Ridall, A.L.; Levine, M.J.  
Biochem. Biophys. Res. Commun. 149, 784-790, 1987  
A>Title: Human submandibular gland statherin and basic histidine-rich peptide are encode  
A:Reference number: A27489; MUID:88106506; PMID:3426601  
A:Accession: B27489  
A:Molecule type: mRNA  
A:Residues: 1-62 <DIC>  
A:Cross-references: GB:M18371; NID:G338610; PIDN:AAA60600.1; PID:G338611  
R:Schlesinger, D.H.; Hay, D.I.  
J. Biol. Chem. 252, 1689-1695, 1977  
A>Title: Complete covalent structure of statherin, a tyrosine-rich acidic peptide which  
A:Reference number: A03288; MUID:77118656; PMID:838735  
A:Accession: A03288  
A:Molecule type: protein  
A:Residues: 20-62 <SCH>  
R:Oppenheim, F.G.; Hay, D.I.; Smith, D.J.; Offner, G.D.; Troxler, R.F.  
J. Dent. Res. 66, 462-466, 1987  
A>Title: Molecular basis of salivary proline-rich protein and peptide synthesis: cell-fir  
gnal peptides.  
A:Reference number: A92773; MUID:87309161; PMID:3476566  
A:Accession: A32524  
A:Molecule type: protein  
A:Residues: 1,'X',3-4,'X',6,'X',8,'X',10,'X',12-13,'XX',16 <OPP>  
C:Note: radiosequencing of precursor after cell-free translation  
A:Comment: Statherin is one of the salivary proteins that stabilize saliva supersaturate  
bly being precursors of enamel pellicle). These inhibitors thus promote enamel stabiliza  
C:Genes: GDB:STATH  
A:Gene: GDB:STATH  
A:Cross-references: GDB:120391; OMIM:184470  
A:Map position: 4q11-4q13

A;Introns: 17/3; 24/3; 34/3  
C;Superfamily: scatherin precursor; scatherin/histatin signal sequence homology  
C;Keywords: phosphoprotein; saliva  
F;1-25/Domain: scatherin/histatin signal sequence homology <SHH>  
F;1-19/Domain: signal sequence #status experimental <SIG>  
F;20-62/Product: scatherin #status experimental <MA1>  
F;21,22/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 34.8%; Score 110; DB 1; Length 62;  
Best Local Similarity 49.1%; Pred. No. 7.5e-06;  
Matches 27; Conservative 7; Mismatches 19; Indels 2; Gaps 2;

QY 1 MKFLVFVFLALMIRADSSSEKRRKRRKHHGFGQYQYQYQYPLNYPAY 55  
Db 1 MKFLVFVFLALMIRADSSSEKRRKRRKHHGFGQYQYQYPLNYPAY 53

RESULT 2  
A32541  
histatin 1 precursor [validated] - human  
N;Alternate names: histidine-rich protein 1  
C;Species: Homo sapiens (man)  
C;Keywords: histatin 2  
F;1-25/Domain: histatin precursor; scatherin/histatin signal sequence homology <SHH>  
F;1-19/Domain: signal sequence #status experimental <SIG>  
F;20-57/Product: histatin 2 #status experimental <MA2>  
F;21/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 32.3%; Score 102; DB 1; Length 57;  
Best Local Similarity 37.7%; Pred. No. 5.7e-05;  
Matches 20; Conservative 13; Mismatches 18; Indels 2; Gaps 1;

QY 1 MKFLVFVFLALMIRADSSSEKRRKRRKHHGFGQYQYQYPLNYPAY 51  
Db 1 MKFLVFVFLALMIRADSSSEKRRKRRKHHGFGQYQYQYPLNYPAY 53

RESULT 3  
B32541  
histatin 3 precursor [validated] - human  
N;Alternate names: histidine-rich protein 3  
C;Species: Homo sapiens (man)  
C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text change 08-Dec-2000  
C;Accession: B32541; A27489; B60659; B60664; C60664; B28164; C28164; B60742; C60742; D60742; E60742; F60742; G60742; H60742; I60742; J60742; K60742; L60742; M60742; N60742; O60742; P60742; Q60742; R60742; S60742; T60742; U60742; V60742; W60742; X60742; Y60742; Z60742

A;Reference number: A60742; MUID:90154430; PMID:2303595  
A;Accession: A60742  
A;Molecule type: protein  
A;Residues: 31-57 <TRO>  
R;Vanderspek, J.C.; Offner, G.D.; Troxler, R.F.; Oppenheim, F.G.  
Arch. Oral Biol. 35, 137-143, 1990  
A;Title: Molecular cloning of human submandibular histatins.  
A;Reference number: A60659; MUID:90262442; PMID:2344289  
A;Accession: A60659  
A;Molecule type: mRNA  
A;Residues: VML, 14-57 <VA2>  
C;Genes: GDB:HTN1  
A;Cross-references: GDB:120068; OMIM:142701  
A;Map position: 4q13-4q13  
A;Introns: 17/3; 24/3; 34/3  
C;Superfamily: histatin precursor; scatherin/histatin signal sequence homology  
C;Keywords: phosphoprotein; saliva  
F;1-25/Domain: scatherin/histatin signal sequence homology <SHH>  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;20-57/Product: histatin 2 #status experimental <MA1>  
F;21/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 32.3%; Score 102; DB 1; Length 57;  
Best Local Similarity 37.7%; Pred. No. 5.7e-05;  
Matches 20; Conservative 13; Mismatches 18; Indels 2; Gaps 1;

QY 1 MKFLVFVFLALMIRADSSSEKRRKRRKHHGFGQYQYQYPLNYPAY 51  
Db 1 MKFLVFVFLALMIRADSSSEKRRKRRKHHGFGQYQYQYPLNYPAY 53

RESULT 3  
B32541  
histatin 3 precursor [validated] - human  
N;Alternate names: histidine-rich protein 3  
C;Species: Homo sapiens (man)  
C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text change 08-Dec-2000  
C;Accession: B32541; A27489; B60659; B60664; C60664; B28164; C28164; B60742; C60742; D60742; E60742; F60742; G60742; H60742; I60742; J60742; K60742; L60742; M60742; N60742; O60742; P60742; Q60742; R60742; S60742; T60742; U60742; V60742; W60742; X60742; Y60742; Z60742

A.Molecule type: protein  
A.Residues: 20-51 <SU2>  
R.Oppenheim, F.G.; Xu, T.; McMillian, F.M.; Levitz, S.M.; Diamond, R.D.; Offner, G.D.; T  
J. Biol. Chem. 263, 7472-7477, 1988  
A.Title: Histatins, a novel family of histidine-rich proteins in human parotid secretion  
A.Reference number: A9685; MUID:88227937; PMID:3286634  
A.Molecule type: protein  
A.Residues: 20-51 <OPP>  
A.Accession: C28164  
A.Molecule type: protein  
A.Residues: 20-43 <OP2>  
R.Troxler, R.F.; Offner, G.D.; Xu, T.; Vanderepek, J.C.; Oppenheim, F.G.  
J. Dent. Res. 69, 2-6, 1990  
A.Title: Structural relationship between human salivary histatins.  
A.Reference number: A60742; MUID:90154430; PMID:2303595  
A.Accession: B60742  
A.Molecule type: protein  
A.Residues: 31-51 <TRO>  
A.Accession: C60742  
A.Molecule type: protein  
A.Residues: 20-44 <TR2>  
A.Accession: D60742  
A.Molecule type: protein  
A.Residues: 31-43 <TR3>  
A.Accession: H60742  
A.Molecule type: protein  
A.Residues: 24-31 <TR7>  
R.Xu, L.; Fischer, T.; Pollock, J.J.  
Pept. Res. 2, 373-375, 1989  
A.Title: Sequence determination of low molecular weight salivary histidine-rich polypept  
A.Reference number: A60675; MUID:92273931; PMID:2520776  
A.Accession: A60675  
A.Molecule type: protein  
A.Residues: 25-43 <XUA>  
R.Mutakant, Y.; Amano, A.; Takagaki, M.; Shizukunishi, S.; Tsunemitsu, A.; Aimoto, S.  
FEMS Microbiol. Lett. 72, 275-280, 1990  
A.Title: Purification and characterization from human parotid secretion of a peptide whi  
A.Reference number: A61000  
A.Accession: A61000  
A.Molecule type: protein  
A.Residues: 32-43 <MUR>  
A.Note: This salivary histidine-rich peptide was shown to inhibit hemagglutination by Ba  
R.Sabatini, L.M.; Ota, T.; Azen, E.A.  
Mol. Biol. Evol. 10, 497-511, 1993  
A.Title: Nucleotide sequence analysis of the human salivary protein genes H1S1 and H1S2,  
A.Reference number: 157425; MUID:93330039; PMID:8336540  
A.Accession: 177213  
A.Status: translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-51 <RES>  
A.Cross-references: GB:I05514; NID:G184056; PIDN:AAA02746.1; PID:G184058  
A.Gene: GDB:HTN3  
C:Genetics:  
A:Cross-references: GDB:125601; OMIM:142702  
A:Map position: 4q12-4q21  
A.Introns: 17/3; 24/3; 34/3  
C:Superfamily: histatin precursor; statherin/histatin signal sequence homology  
C:Keywords: saliva  
F:1-25/Domains: statherin/histatin signal sequence homology <SHR>  
F:1-19/Domains: signal sequence #status predicted <SIG>  
F:20-51/Product: histatin 3 #status experimental <MA3>  
F:20-44/Product: histatin 6 #status experimental <MA5>  
F:20-43/Product: histatin 5 #status experimental <MA1>  
F:24-31/Product: histatin 11 #status experimental <M11>  
F:24-30/Product: histatin 12 #status experimental <M12>  
F:31-51/Product: histatin 4 #status experimental <MA4>  
F:31-44/Product: histatin 9 #status experimental <MA9>  
F:31-43/Product: histatin 7 #status experimental <MA7>  
F:32-44/Product: histatin 10 #status experimental <M10>  
F:32-43/Product: histatin 8 #status experimental <MA8>  
Query Match 30.1%; Score 95; DB 1; Length 51;

Best Local Similarity 46.8%; Pred. No. 0.00033;  
Matches 22; Conservative 6; Mismatches 13; Indels 6; Gaps 2;  
Qy 1 MKRIFVFMILIMLRADSSSEKRRKRRK-----HGYPQY 41  
Db 1 MKFVFAILIALMTGADSHARKHHQYKRRFHKHSHRGYRSNY 47  
RESULT 4  
SMMQPI  
statherin precursor - crab-eating macaque  
C:Species: Macaca fascicularis (Crab-eating macaque)  
C:Dates: 17-Dec-1982 #sequence\_revision 21-Jul-1995 #text\_change 21-Jan-2000  
R:Accession: B32524; A03289  
R:Oppenheim, F.G.; Hay, D.I.; Smith, D.J.; Offner, G.D.; Troxler, R.F.  
J. Dent. Res. 66, 462-466, 1987  
A.Title: Molecular basis of salivary proline-rich protein and peptide synthesis: cell-f  
gnal peptides.  
A.Reference number: A92773; MUID:87309161; PMID:3476566  
A.Accession: B32524  
A.Molecule type: protein  
A.Residues: 1-19 <OP1>  
A.Note: 6-Leu was also found  
R.Oppenheim, F.G.; Offner, G.D.; Troxler, R.F.  
J. Biol. Chem. 257, 9271-9282, 1982  
A.Title: Phosphoproteins in the parotid saliva from the subhuman primate Macaca fascicul  
of a proline-rich phosphopeptide.  
A.Reference number: A03289; MUID:82265555; PMID:7107568  
A.Accession: A03289  
A.Molecule type: protein  
A.Residues: 20-61 <OP2>  
C:Superfamily: statherin precursor; statherin/histatin signal sequence homology  
C:Keywords: phosphoprotein; saliva  
F:1-25/Domains: statherin/histatin signal sequence homology <SHR>  
F:1-19/Domains: signal sequence #status experimental <SIG>  
F:20-61/Product: statherin #status experimental <MA7>  
F:21-22/Binding site: phosphate (Ser) (covalent) #status experimental  
Query Match 25.8%; Score 81.5; DB 1; Length 61;  
Best Local Similarity 34.5%; Pred. No. 0.014;  
Matches 19; Conservative 7; Mismatches 28; Indels 1; Gaps 1;  
Qy 1 MKRIFVFMILIMLRADSSSEKRRKRRKGGYQOYQYQYPLNPPY 55  
Db 1 MKFLKFLXLLXKXKXSSSEKFLRRFRPDEGRYQYQYQYPL 54  
RESULT 5  
JCS233  
Spermatophorin Sp23 - yellow mealworm  
C:Species: Tribolium molitor (yellow mealworm)  
C:Dates: 13-Mar-1997 #sequence\_revision 13-Mar-1997 #text\_change 21-Jul-2000  
R:Accession: JCS233  
R:Feng, X.; Happ, G.M.  
Gene 179, 257-262, 1996  
A.Title: Isolation and sequencing of the gene encoding Sp23, a structural protein of spe  
A.Reference number: JCS233; MUID:97128316; PMID:8972909  
A.Accession: JCS233  
A>Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-181 <FEN>  
A:Cross-references: GB:U9658; NID:G1079704; PIDN:AA047422.1; PID:G1079705  
A>Note: the authors translated the codon CAA for residue 149, and CAG for residues 24, 5  
C:Comment: This protein is a structural component that is mapped to a layer within the w  
C:Genetics:  
A.Gene: Sp23  
A.Introns: 7/1  
C:Superfamily: hydroxyproline-rich glycoprotein  
Query Match 23.9%; Score 75.5; DB 2; Length 181;  
Best Local Similarity 31.2%; Pred. No. 0.18;  
Matches 15; Conservative 12; Mismatches 14; Indels 7; Gaps 2;





C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 11-May-2000  
 C/Accession: T06211; A03355  
 R.Sainova, O.V.; Mekhedov, S.L.; Zhelnin, L.G.; Khokhlova, T.A.; Anan'ev, E.V.  
 Genetika 29, 1070-1079, 1993  
 A>Title: Nucleotide sequence of the barley C-hordein gene.  
 A/Reference number: Z15537; MUID:93380629; PMID:8396543  
 A/Accession: T06211  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-310 <SAI>  
 A/Cross-references: EMBL:566938; NID:g442523; PIDN:AB28161.1; PID:g442524  
 R.Sherry, P.R.; Autran, J.C.; Nimmo, C.C.; Lew, E.J.U.; Kaserda, D.D.  
 Nature 286, 520-522, 1980  
 A/Reference number: A93228  
 A/Accession: A03355  
 A/Molecule type: protein  
 A/Residues: 21-48 <SHE>  
 C/Superfamily: gliadin  
 F.1-20/Domains: signal sequence #status predicted <SIG>  
 F.21-310/Product: C-hordein #status experimental <MAT>

Query Match 20.9%; Score 66; DB 2; Length 310;  
 Best Local Similarity 25.9%; Pred. No. 3.6; Indels 4; Gaps 2;  
 Matches 15; Conservative 14; Mismatches 25;

Qy 1 MKIFVFVIMALIMIRADSSSEKRRKRGVFOQYQYRPLNPPAYPP 58  
 1 MKIFVFVIMALIMIRADSSSEKRRKRGVFOQYQYRPLNPPAYPP 54

## RESULT 11

T39588  
 Hypothetical protein SPBC16E9.14c - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 R.Volckaert, G.; Wood, V.; Rajandream, M.A.; Bartell, B.G.  
 submitted to the EMBL Data Library, August 1997  
 A/Reference number: Z21865  
 A/Accession: T39588  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-366 <VOL>  
 A/Cross-references: EMBL:299759; PIDN:CA16906.1; GSPDB:GN00067; SPDB:SPBC16E9.14c  
 A/Experimental source: strain 97zh-; cosmid c16E9  
 C/Genetics:  
 A/Gene: SPDB:SPBC16E9.14c  
 A/Map position: 2

Query Match 20.9%; Score 66; DB 2; Length 386;  
 Best Local Similarity 36.6%; Pred. No. 4.4; Indels 4; Gaps 1;  
 Matches 15; Conservative 4; Mismatches 18;

Qy 22 SEKKRRKRRKRGVFOQYQYRPLNPPAYPP 58  
 41 SVDTFRTRKRRKHHKSHSHQYFLPKRNPPLPASPYP 81

## RESULT 12

T51434  
 Hypothetical protein F2G14.10 - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
 R.Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew  
 submitted to the Protein Sequence Database, August 2000  
 A/Reference number: Z25394  
 A/Accession: T51434  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-733 <SAT>  
 A/Cross-references: EMBL:AL391146  
 A/Experimental source: cultivar Columbia; BAC clone F2G14

C/Genetics:  
 A/Map position: 5  
 A/Accession: 44/1; 104/1; 161/1; 230/1; 260/3; 459/3; 499/3  
 A/Note: F2G14\_10

Query Match 20.9%; Score 66; DB 2; Length 733;  
 Best Local Similarity 32.6%; Pred. No. 7.9; Indels 0; Gaps 0;  
 Matches 14; Conservative 8; Mismatches 21;

Qy 6 FVFIMALIMIRADSSSEKRRKRGVFOQYQYRPLNPPAYPP 48  
 Db 240 FGVMALIVRRVSLFSSSHDTKSRHVATPMTAPYQRYP 282

## RESULT 13

F82862  
 conjugal transfer protein Xfa0014 [imported] - Xylella fastidiosa (strain 9a5c)  
 C/Species: Xylella fastidiosa  
 C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C/Accession: F82862  
 R.anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences  
 Nature 406, 151-157, 2000  
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A/Reference number: A82515; MUID:20365717; PMID:10910347  
 A/Note: for a complete list of authors see reference number A59328 below  
 A/Accession: F82862  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-447 <SIM>

A/Cross-references: GB:AB003851; NID:g9112238; PIDN:AF85583.1; GSPDB:GN00130; XFSC:XFA0  
 A/Experimental source: strain 9a5c  
 R.Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
 as-Neco, E.; Docena, C.; El-Dorri, H.; Facinani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000

A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitejima, J.P.; Kuramae, E.E.; Laig  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martins, C.L.; Marques, M.V.; Martins, E  
 A/Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeiri, D.A  
 Rodrigues, V.; Rosa, A.C.R.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tenuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A/Reference number: A59328  
 A/Contents: annotation  
 C/Genetics:  
 A/Gene: XFA0014  
 A/Genome: plasmid  
 A/Note: plasmid pXFS.1

Query Match 20.7%; Score 65.5; DB 2; Length 447;  
 Best Local Similarity 32.3%; Pred. No. 5.8; Indels 15; Gaps 3;  
 Matches 20; Conservative 12; Mismatches 15;

Qy 3 IFIFVIMALIMIRADSSSEKRRKRGVFOQYQYRPLNPPAYPP 56  
 Db 76 IFGFIFLSAAMVAVNGDKKKKKKKK-----QPDQKQGIANNLPPLVTPAAP 126

Qy 57 FP 58  
 Db 127 AP 128

## RESULT 14

T21275  
 Hypothetical protein F22B12.1 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 R.Matthews, P.  
 submitted to the EMBL Data Library, April 1996  
 A/Reference number: Z19399  
 A/Accession: T21275

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1203 <MIL>

A;Cross-references: EMBL:Z71180; PIDN:CAA94892.1; GSPDB:GN00023; CESP:F22E12.1

A;Experimental source: clone F22E12

C;Genetics:

A;Gene: CESP:F22E12.1

A;Map position: 5

A;Introns: 71/1; 156/3; 223/3; 339/1; 669/3; 736/1; 758/3; 854/3; 931/3; 968/3; 1181/3

#### Query Match

20.7%; Score 65.5; DB 2; Length 1203;

Best Local Similarity 38.5%; Pred. No. 14;

Matches 15; Conservative 7; Mismatches 16; Indels 1; Gaps 1;

QY 20 DSSEKRRKKKHHGYYFOQYQYQYPLNYPPAYPPP 58

DB 354 NEAEETQRIQDTMRNTEQAR-YQRPVQMPPPPPP 391

#### RESULT 15

S18350

Seed storage protein - barley

C;Species: Hordeum vulgare (barley)

C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Aug-1999

C;Accession: S18350

R;Entwistle, J.; Knudsen, S.; Mueller, M.; Cameron-Mills, V.

Plant Mol. Biol. 17, 1217-1231, 1991

A;Title: Amber codon suppression: the in vivo and in vitro analysis of two C-hordein gen

A;Reference number: S18350; MUID:92032786; PMID:1932695

A;Accession: S18350

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-260 <ENT>

A;Cross-references: EMBL:X60037; NID:G19000; PIDN:CAA42642.1; PID:G19001

C;Superfamily: gliadin

#### Query Match

20.6%; Score 65; DB 2; Length 260;

Best Local Similarity 25.9%; Pred. No. 4;

Matches 15; Conservative 14; Mismatches 25; Indels 4; Gaps 2;

QY 1 MKTIFVFMALIAMIRADSSSEKRRKKHKGYYFOQYQYQYPLNYPPAYPPP 58

DB 1 MKTFPLFVLAMVMSIVTARQLNPSSQELQSPQSYLQ-QPYQNP--YLPQKPP 54

Search completed: December 13, 2003, 03:09:11  
Job time : 26 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 13, 2003, 02:29:24 ; Search time 16 Seconds  
(without alignments)  
170.472 Million cell updates/sec

Title: US-10-079-754A-10

Perfect score: 316  
Sequence: 1 MKIFIFVFMALITAMRAD.....OOYCPYGRYPINYPAPPFP 58

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	34.8	62	1 STAT_HUMAN	P02808 homo sapien
2	102	32.3	57	1 HIS1_HUMAN	P15515 homo sapien
3	95	30.1	51	1 HIS3_HUMAN	P15516 homo sapien
4	81.5	25.8	61	1 STAT_MACFA	P02809 macaca fasc
5	75.5	23.9	182	1 SP23_TENMO	O27022 tenebrio mo
6	70.5	22.3	42	1 STAT_MACAR	P14709 macaca arcu
7	66.5	21.0	180	1 CASK_RABIT	P33618 oryctolagus
8	64.5	20.7	1362	1 BRD4_HUMAN	O60885 homo sapien
9	64.5	20.4	533	1 YD3D_SCHPO	Q10277 schizosach
10	63.5	20.1	293	1 HOR1_HORVU	P06470 hordeum vul
11	63	19.9	182	1 CASK_HUMAN	P07498 homo sapien
12	61	19.3	377	1 CC37_TETRL	O9d9q7 tetraodon f
13	60.5	19.1	226	1 CASK_HUMAN	P05814 homo sapien
14	60	19.0	1215	1 C1XA_BACTM	O45715 bacillus th
15	59	18.7	409	1 MOBA_THIFE	P22898 thiochacilli
16	58.5	18.5	2142	1 BAT2_HUMAN	P48634 homo sapien
17	58.5	18.5	193	1 VIR1_AEPAE	O16927 aedes aegypt
18	58	18.4	133	1 CAS2_CAMDR	O97944 camelus dro
19	57.5	18.2	234	1 IRI1_HOMVA	P16809 human cytom
20	57.5	18.2	404	1 GYRB_BACCE	O9x3y3 bacillus ce
21	57	18.0	455	1 YP46_YEAST	Q12080 saccharomyc
22	57	18.0	698	1 PPK_XYLFA	O9pac7 xyliella fas
23	57	18.0	1073	1 HRJ3_DROME	P48669 drosophila
24	57	18.0	1097	1 KRID_RAT	O35877 rattus norv
25	56.5	17.9	1290	1 PER1_HUMAN	O15534 homo sapien
26	56	17.7	281	1 BLAC_STAUV	P00807 staphylococ
27	56	17.7	636	1 NAF1_HUMAN	Q15025 homo sapien
28	56	17.7	861	1 POS8_CAEEL	P34552 caenorhabdi
29	56	17.7	873	1 RX_DROME	O9w4g1 drosophila
30	56	17.7	1294	1 RRO_MCMVM	P09398 white clove
31	56	17.7	1294	1 RRO_MCMVM	P15402 white clove
32	55.5	17.6	321	1 YBI_CHICK	Q06066 gallus gall
33	55.5	17.6	556	1 WSC3_YEAST	Q12215 saccharomyc

34	55	17.4	308	1 MGFI_HUMAN	O9hay2 homo sapien
35	55	17.4	470	1 NRS4_HUMAN	Q15233 homo sapien
36	55	17.4	670	1 NS21_MOUSE	Q91yvs mus musculu
37	54.5	17.2	282	1 GDA3_THEAT	P04723 triticum ae
38	54.5	17.2	456	1 MGAI_YEAST	P53050 saccharomyc
39	54.5	17.2	971	1 CL44_CANAL	O14427 candida alb
40	54.5	17.2	4590	1 FATH_HUMAN	O14517 homo sapien
41	54	17.1	137	1 SMR2_RAT	P18897 rattus norv
42	54	17.1	178	1 TRSF_DROER	O23935 drosophila
43	54	17.1	184	1 TRSF_DROCI	Q24669 drosophila
44	54	17.1	190	1 CASK_BUBBU	P11840 bubalus bub
45	54	17.1	197	1 TRSF_DROME	P11596 drosophila

## ALIGNMENTS

RESULT 1

ID	STAT_HUMAN	STANDARD	PRT	62 AA.
AC	P02808:			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Statherin precursor.			
GN	STAT.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88074310; PubMed=3502720;			
RA	Sabatini L., Carllock L., Johnson G., Azen E.;			
RT	"CDNA cloning and chromosomal localization (4q11-13) of a gene for			
RT	statherin, a regulator of calcium in saliva.";			
RL	Am. J. Hum. Genet. 41:1048-1060(1987).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88106506; PubMed=3426601;			
RA	Dickinson D.P., Ridall A.L., Levine M.J.;			
RT	"Human submandibular gland statherin and basic histidine-rich peptide			
RT	are encoded by highly abundant mRNA's derived from a common ancestral			
RT	sequence.";			
RL	Biochem. Biophys. Res. Commun. 149:784-790(1987).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90323623; PubMed=2373369;			
RA	Sabatini L.M., He Y.-Z., Azen E.A.;			
RT	"Structure and sequence determination of the gene encoding human			
RT	salivary statherin.";			
RL	Gene 89:245-251(1990).			
RN	[4]			
RP	SEQUENCE OF 20-62.			
RX	MEDLINE=77118656; PubMed=838735;			
RA	Schlesinger D.H., Hay D.I.;			
RT	"Complete covalent structure of statherin, a tyrosine-rich acidic			
RT	peptide which inhibits calcium phosphate precipitation from human			
RT	parotid saliva.";			
RL	J. Biol. Chem. 252:1689-1695(1977).			
RN	[5]			
RP	PARTIAL SEQUENCE OF 1-16.			
RX	MEDLINE=87309161; PubMed=3476566;			
RA	Oppenheim F.G., Hay D.I., Smith D.J., Offner G.D., Troxler R.F.;			
RT	"Molecular basis of salivary proline-rich protein and peptide			
RT	synthesis: cell-free translations and processing of human and macaque			
RT	statherin mRNAs and partial amino acid sequence of their signal			
RT	peptides.";			
RL	J. Dent. Res. 66:462-466(1987).			
CC	-I- FUNCTION: SALIVARY PROTEIN THAT STABILIZES SALIVA SUPERSATURATED			
CC	WITH CALCIUM SALTS BY INHIBITING THE PRECIPITATION OF CALCIUM			
CC	PHOSPHATE SALTS. IT ALSO MODULATES HYDROXYAPATITE CRYSTAL			
CC	FORMATION ON THE TOOTH SURFACE.			

RT "Nucleotide sequence analysis of the human salivary protein genes  
RT H1S1 and H1S2, and evolution of the STRATH/HIS gene family.",  
RL Mol. Biol. Evol. 10:497-511(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skeleral muscle;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Tohiyuki S., Canninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Wolley P.C., Hale S., Garcia A.M., Gay L.U., Huliy S.W.,  
RA Valladao D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S., Sanchez  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.C., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield A.J., Schein J.E., Jones S.J.M., Marra M.A.,  
RA Schnerch A., Sehnin J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.",  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP SEQUENCE OF 20-57.  
RC TISSUE=Parotid gland;  
RX MEDLINE=88227937; PubMed=3286634.  
RA Oppenheim F.G., Xu T., McMillan F.M., Levitz S.M., Diamond R.D.,  
RA Offner G.D., Troxler R.F.,  
RT "Histatins, a novel family of histidine-rich proteins in human  
RT parotid secretion. Isolation, characterization, primary structure,  
RT and functional effects on *Candida albicans*.",  
RL J. Biol. Chem. 263:7472-7477(1998).  
RN [5]  
RP SEQUENCE OF 20-57.  
RC TISSUE=Parotid gland;  
RX MEDLINE=8611175; PubMed=3944083;  
RA Oppenheim F.G., Yang Y.C., Diamond R.D., Hyslop D., Offner G.D.,  
RA Troxler R.F.,  
RT "The primary structure and functional characterization of the neutral  
RT histidine-rich polypeptide from human parotid secretion.",  
RL J. Biol. Chem. 261:1177-1182(1986).  
RN [6]  
RP SEQUENCE OF 20-57.  
RC TISSUE=Saliva;  
RX MEDLINE=90321151; PubMed=2372245;  
RA Sugiyama K., Ogino T., Ogata K.,  
RT "Rapid purification and characterization of histatine (histidine-rich  
RT polypeptides) from human saliva.",  
RL Arch. Oral Biol. 35:415-419(1990).  
RN [7]  
RP SEQUENCE OF 14-57 FROM N.A.  
RX MEDLINE=89371745; PubMed=2773933;  
RA Vanderspek J.C., Wyandt H.E., Skare J.C., Milunsky A., Oppenheim F.G.,  
RA Troxler R.F.,  
RT "Localization of the genes for histatins to human chromosome 4q13 and  
RT tissue distribution of the mRNA.",  
RL Am. J. Hum. Genet. 45:381-387(1989).  
CC -1- FUNCTION: HISTATINS ARE SALIVARY PROTEINS THAT ARE CONSIDERED TO  
CC BE MAJOR PRECURSORS OF THE PROTECTIVE PROTEINOUS STRUCTURE ON  
CC TOOTH SURFACES (ENAMEL PELlicLE). IN ADDITION, HISTATINS EXHIBIT  
CC ANTIBACTERIAL AND ANTIFUNGAL ACTIVITIES.  
CC -1- PPM: HISTATIN 2 IS A PROTEOLYTIC PRODUCT OF HISTATIN 1.  
CC -1- SIMILARITY: TO HISTATIN 3.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; M26664; AAA58645.1; -  
DR EMBL; L04132; AAA02746.1; -  
DR EMBL; BC017835; AAH17835.1; -  
DR PIR; I57425; A32541.  
DR Genew; HGNC:5283; HTN1.  
DR MIM; 142701; -  
DR GO; GO:0005576; Cerebral cellular; NAS.  
DR GO; GO:0003797; F-antibacterial peptide activity; NAS.  
DR GO; GO:0006952; P-defense response; NAS.  
DR Biomat; Biomat; Fungicide; Antibiotic; Signal; Phosphorylation.  
FT SIGNAL 1 19  
FT PEPTIDE 20 57 HISTATIN 1.  
FT MOD\_RES 31 57 HISTATIN 2.  
FT MOD\_RES 21 21 PHOSPHORYLATION.  
SQ SEQUENCE 57 AA; 6963 MW; F532HDIDCE23D83 CRC64;

Query Match 32.3%; Score 102; DB 1; Length 57;  
Best Local Similarity 37.7%; Pred. No. 4.4e-05;  
Matches 20; Conservative 13; Mismatches 18; Indels 2; Gaps 1;

OY 1 MKIFIFPIMILIMIRADSESEKRRKRRKRGY--FOGYOPYORYPLNY 51  
DB 1 MKFVFALVLMISMISDSHEKHHGRRKFKHSHRFPYGYGSNY 53

# RESULT 3

HIS3\_HUMAN STANDARD; PRT; 51 AA.

AC P1516; O16243;  
DT 01-APR-1990 (Rel. 14. Created)  
DT 01-FEB-1991 (Rel. 17. Last sequence update)  
DT 15-SEP-2003 (Rel. 42. Last annotation update)  
DE Histatin 3 precursor (Histidine-rich protein 3) (PB) (Basic histidine-  
rich protein) [Contains: Histatins 4 to 12].  
GN HTN3 OR HIS2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89246491; PubMed=2719677;  
RA Sabatini L.M., Azen E.A.;  
RT "Histatins, a family of salivary histidine-rich proteins, are encoded  
by at least two loci (HIS1 and HIS2).";  
RL Biochem. Biophys. Res. Commun. 160:495-502 (1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88106506; PubMed=3426601;  
RA Dickinson D.P., Ridall A.L., Levine M.J.;  
RT "Human submandibular gland statherin and basic histidine-rich peptide  
are encoded by highly abundant mRNA's derived from a common ancestral  
sequence";  
RL Biochem. Biophys. Res. Commun. 149:784-790 (1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90262442; PubMed=2344289;  
RA Vanderspek J.C., Offner G.D., Troxler R.F., Oppenheim F.G.;  
RT "Molecular cloning of human submandibular histatins";  
RL Arch. Oral Biol. 35:137-143 (1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93330039; PubMed=8336540;  
RA Chen Z.W.;  
RT "Nucleotide sequence analysis of the human salivary protein genes  
HIS1 and HIS2, and evolution of the STATH/HIS gene family.";  
RL Mol. Biol. Evol. 10:497-511 (1993).  
RN [5]  
RP SEQUENCE FROM N.A.

RC TISSUE=Skeletal muscle;  
RX MEDLINE=23388257; PubMed=12477932;  
RA Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshlynski S., Carrinoli P., Prange C.,  
RA Raha S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Boek S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Mazny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [6]  
RP SEQUENCE OF 20-51 FROM N.A.  
RC TISSUE=Saliva;  
RX MEDLINE=95038764; PubMed=7951254;  
RA Sabatini L.M., Azen E.A.;  
RT "Two coding change mutations in the HIS2(2) allele characterize the  
salivary histatin 3-2 protein variant";  
RL Hum. Mutat. 4:12-19 (1994).  
RN [7]  
RP SEQUENCE OF 20-51.  
RC TISSUE=Parotid gland;  
RX MEDLINE=88227937; PubMed=3286634;  
RA Oppenheim F.G., Xu T., McMillan F.M., Levitz S.M., Diamond R.D.,  
RA Offner G.D., Troxler R.F.;  
RT "Histatine, a novel family of histidine-rich proteins in human  
parotid secretion. Isolation, characterization, primary structure,  
and fungistatic effects on Candida albicans";  
RL J. Biol. Chem. 263:7472-7477 (1988).  
RN [8]  
RP SEQUENCE OF 20-42.  
RC TISSUE=Saliva;  
RX MEDLINE=90321151; PubMed=2372245;  
RA Sugiyama K., Ogino T., Ogata K.;  
RT "Rapid purification and characterization of histatins (histidine-rich  
polypeptides) from human whole saliva";  
RL Arch. Oral Biol. 35:415-419 (1990).  
CC -!- FUNCTION: HISTATINS ARE SALIVARY PROTEINS THAT ARE CONSIDERED TO  
BE MAJOR PRECURSORS OF THE PROTECTIVE PROTEINACIOUS STRUCTURE ON  
TOOTH SURFACES (ENAMEL PELICLE). IN ADDITION, HISTATINS EXHIBIT  
ANTIBACTERIAL AND ANTIFUNGAL ACTIVITIES.  
CC -!- PFM: HISTATINS 4 TO 12 ARE PROTEOLYTIC PRODUCTS OF HISTATIN 3.  
CC -!- POLYMORPHISM: THERE ARE TWO ALLELES OF HIS2, HIS2(1) (SHOWN HERE)  
AND HIS2(2) THAT CODES FOR THE VARIANT HISTATIN 3-2 FOUND  
PRIMAIRLY AND IN HIGH FREQUENCIES IN BLACK POPULATIONS.  
CC -!- SIMILARITY: TO HISTATIN 1.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; M26665; AAA58646.1; -  
DR EMBL; M18372; AAA51830.1; -  
DR EMBL; L05514; AAA02746.1; -  
DR EMBL; L05513; AAA02746.1; JOINED.  
DR EMBL; BC009791; AAH09791.1; -  
DR EMBL; S74382; AAB32411.1; -  
DR PIR; B32541; B32541.

DR Genew; HGNC:5284; HTN3.  
 DR MIM; 142702; -  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003797; F:antibacterial peptide activity; NAS.  
 DR GO; GO:0006952; P:defense response; NAS.  
 DR GO; GO:0003045; F:structural constituent of tooth enamel; ISS.  
 DR GO; GO:0030502; P:negative regulation of bone mineralization; ISS.  
 DR GO; GO:0046541; P:saliva secretion; ISS.  
 DR InterPro: IPR005575; Statherin.  
 DR Pfam; PF03875; Statherin; 1.  
 DR Biomimetalization; Phosphorylation; Signal.  
 KM Biomimetalization; Signal.  
 FT SIGNAL 1 19  
 FT PEPTIDE 20 51 HISTATIN 3.  
 FT PEPTIDE 20 43 HISTATIN 5.  
 FT PEPTIDE 20 44 HISTATIN 6.  
 FT PEPTIDE 24 30 HISTATIN 12.  
 FT PEPTIDE 24 31 HISTATIN 11.  
 FT PEPTIDE 31 43 HISTATIN 7.  
 FT PEPTIDE 31 44 HISTATIN 9.  
 FT PEPTIDE 31 51 HISTATIN 4.  
 FT PEPTIDE 32 43 HISTATIN 8.  
 FT PEPTIDE 32 44 HISTATIN 10.  
 FT PEPTIDE 41 41 HISTATIN 10.  
 FT VARIANT 41 41 R -> Q (IN HISTATIN 3-2; LOSS OF THE PROTEOLYTIC CLEAVAGE SITE).  
 FT MISSING (IN HISTATIN 3-2).  
 FT VARIANT 47 51 /FTID=VAR\_005289.  
 FT SEQUENCE 51 AA; 6149 MW; AFCCB4B32083FD65 CRC64;  
 SQ  
 Query Match 30.1%; Score 95; DB 1; Length 51;  
 Best Local Similarity 46.8%; Pred. No. 0.00024;  
 Matches 22; Conservative 6; Mismatches 13; Indels 6; Gaps 2;  
 Oy 1 MKIFIFVFMALIAMIRADSESEKRRKRRKKH---HEGYFOQY 41  
 Db 1 MKFVFVFMALIAMIRADSESEKRRKRRKKHGYKRRKFKHSHRGYSNY 47  
 RESULT 4  
 STAT MACFA STANDARD; PRT; 61 AA.  
 ID STAT MACFA  
 AC P02809;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Statherin precursor.  
 GN STATH.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 NC NCB1\_TaxID=9541;  
 RN [1]  
 RP PARTIAL SEQUENCE OF 1-19.  
 RX MEDLINE=87309161; PubMed=3476566;  
 RA Oppenheim F.G., Hay D.T., Smith D.J., Offner G.D., Troxler R.F.;  
 RT "Molecular basis of salivary proline-rich protein and peptide  
 synthesis: cell-free translations and processing of human and macaque  
 statherin mRNAs and partial amino acid sequence of their signal  
 peptides.";  
 RT J. Dent. Res. 66:462-466(1987).  
 RL [2]  
 RP SEQUENCE OF 20-61.  
 RX MEDLINE=8226555; PubMed=7107568;  
 RA Oppenheim F.G., Offner G.D., Troxler R.F.;  
 RT "Phosphoproteins in the parotid saliva from the subhuman primate  
 Macaca fascicularis. Isolation and characterization of a proline-rich  
 phosphoglycoprotein and the complete covalent structure of a proline-  
 rich phosphopeptide.";  
 RL J. Biol. Chem. 257:9271-9282(1982).  
 CC -1- FUNCTION: SALIVARY PROTEIN THAT STABILIZES SALIVA SUPERSATURATED  
 WITH CALCIUM SALTS BY INHIBITING THE PRECIPITATION OF CALCIUM  
 PHOSPHATE SALTS. IT ALSO MODULATES HYDROXYAPATITE CRYSTAL  
 FORMATION ON THE TOOTH SURFACE.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY PAROTID AND SUBMANDIBULAR  
 GLANDS.

DR PIR; B32524; SEMQPI.  
 DR GO; GO:0005576; C:extracellular; ISS.  
 DR GO; GO:0030197; F:extracellular matrix constituent, lubricant. . .; ISS.  
 DR GO; GO:0046848; F:hydroxyapatite binding activity; ISS.  
 DR GO; GO:0005515; F:protein binding activity; ISS.  
 DR GO; GO:0030345; F:structural constituent of tooth enamel; ISS.  
 DR GO; GO:0030502; P:negative regulation of bone mineralization; ISS.  
 DR GO; GO:0046541; P:saliva secretion; ISS.  
 DR InterPro: IPR005575; Statherin.  
 DR Pfam; PF03875; Statherin; 1.  
 DR Biomimetalization; Phosphorylation; Signal.  
 KM Biomimetalization; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 61 STATHERIN.  
 FT MOD\_RES 21 21 PHOSPHORYLATION.  
 FT MOD\_RES 22 22 PHOSPHORYLATION.  
 FT DOMAIN 20 25 HYDROXYAPATITE-BINDING (INHIBITS  
 CRYSTAL GROWTH).  
 FT DOMAIN 37 61 HYDROPHOBIC (INHIBITS PRECIPITATION  
 OF CALCIUM PHOSPHATE SALTS).  
 FT SEQUENCE 61 AA; 7452 MW; 64241AA8B5641A5B CRC64;  
 SQ  
 Query Match 25.8%; Score 81.5; DB 1; Length 61;  
 Best Local Similarity 34.5%; Pred. No. 0.0092;  
 Matches 19; Conservative 7; Mismatches 28; Indels 1; Gaps 1;  
 Oy 1 MKIFIFVFMALIAMIRADSESEKRRKRRKKHGYFOQYORYPLNYPPAY 55  
 Db 1 MKFLFXPLXLLKXKXKXSDSESEKFLRLRRDRGKRGYQFAPQPL-YRQPY 54  
 RESULT 5  
 SP23 TENMO STANDARD; PRT; 182 AA.  
 ID SP23 TENMO  
 AC Q27022; Q27012;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Spermatophorin SP23 precursor.  
 GN SP23.  
 OS Tenebrio molitor (yellow mealworm).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformata;  
 OC Tenebrionidae; Tenebrio.  
 NC NCB1\_TaxID=7067;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 8-20.  
 RC TISSUE=Bean-shaped accessory gland;  
 RX MEDLINE=92406806; PubMed=1527013;  
 RA Paesen G.C., Schwartz M.B., Peferoen M., Weyda F., Happ G.M.;  
 RT "Amino acid sequence of Sp23, a structural protein of the  
 spermatophore of the mealworm beetle, Tenebrio molitor.";  
 RL J. Biol. Chem. 267:18852-18857(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97128316; PubMed=8972909;  
 RA Feng X., Happ G.M.;  
 RT "Isolation and sequencing of the gene encoding Sp23, a structural  
 protein of spermatophore of the mealworm beetle, Tenebrio molitor.";  
 RL Gene 179:257-262(1996).  
 CC -1- FUNCTION: STRUCTURAL PROTEIN OF A LAYER WITHIN THE WALL OF THE  
 GLAND (BAG). FIXATION IN THE SPERMATOPHORE SEEMS TO REQUIRE  
 COVALENT CROSS-LINKING OF SPERMATOPHORINS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: SPERMATOPHORE.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT THE BEGINNING OF THE ADULT  
 STAGE, ABOUT FOUR DAYS AFTER THE PUPAL PEAK OF ECDYSTEROIDS.  
 CC -1- SIMILARITY: STRUCTURAL RESEMBLANCE WITH D. MELANOGASTER S36 AND  
 S38 CHROMATINIC PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M92928; AAA18165.1; -  
 DR EMBL; U39658; AAC47422.1; -  
 DR PIR; A4157; A4157.  
 DR PIR; JCS233; JCS233.  
 KW Signal; Repeat; Structural protein.  
 FT SIGNAL 1 7  
 FT CHAIN 8 182 SPERMATOPHYTES SP23.  
 FT DOMAIN 110 114 POLY-PRO.  
 FT DOMAIN 120 127 POLY-PRO.  
 FT CONFLICT 61 61 K -> R (IN REF. 2).  
 FT CONFLICT 110 110 P -> A (IN REF. 2).  
 FT CONFLICT 127 127 MISSING (IN REF. 2).  
 SQ SEQUENCE 182 AA; 20570 MW; D84F0F95AC1E3C CRC64;

Query Match 23.9%; Score 75.5; DB 1; Length 182;  
 Best Local Similarity 31.2%; Pred. No. 0.12;  
 Matches 15; Conservative 12; Mismatches 14; Indels 7; Gaps 2;

OY 16 MIRADSEKRRKKHRRGFGQYQ-----PYGRPLNTPPAYPP 58  
 DB 1 MVASIAGEBPAEAKSQSPHFGPYRPPYPPYGRPYTP--YYP 46

RESULT 6  
 STAT\_MACAR STANDARD; PRT; 42 AA.  
 AC P14709;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Statherin.  
 GN STATH.  
 OS Macaca arctoides (Stump-tailed macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9540;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=90129552; PubMed=2613439;  
 RA Schlesinger D.H., Hay D.I., Levine M.J.;  
 RT "Complete primary structure of statherin, a potent inhibitor of  
 RT calcium phosphate precipitation, from the saliva of the monkey,  
 RT Macaca arctoides.";  
 RL Int. J. Pept. Protein Res. 34:374-380(1989).  
 CC -1- FUNCTION: SALIVARY PROTEIN THAT STABILIZES SALIVA SUPERSATURATED  
 CC WITH CALCIUM SALTS BY INHIBITING THE PRECIPITATION OF CALCIUM  
 CC PHOSPHATE SALTS. IT ALSO MODULATES HYDROXYAPATITE CRYSTAL  
 CC FORMATION ON THE TOOTH SURFACE.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY PAROTID AND SUBMANDIBULAR  
 CC GLANDS.  
 CC PIR: JK0224;  
 DR GO; GO:0005576; C:extracellular; ISS.  
 DR GO; GO:0030157; F:extracellular matrix constituent; lubricant. .; ISS.  
 DR GO; GO:0046848; F:hydroxyapatite binding activity; ISS.  
 DR GO; GO:0005515; F:protein binding activity; ISS.  
 DR GO; GO:0030345; F:structural constituent of tooth enamel; ISS.  
 DR GO; GO:0030502; P:negative regulation of bone mineralization; ISS.  
 DR GO; GO:0046541; P:saliva secretion; ISS.  
 DR InterPro; IPR005575; Statherin.  
 DR Pfam; PF03875; Statherin; 1.  
 KW Biomineralization; Phosphorylation.  
 FT MOD\_RES 2 2 PHOSPHORYLATION.  
 FT MOD\_RES 3 3 PHOSPHORYLATION.  
 FT MOD\_RES 1 1 HYDROXYAPATITE-BINDING (INHIBITS  
 FT DOMAIN 18 42 CRYSTAL GROWTH).  
 FT DOMAIN 18 42 HYDROPHOBIC (INHIBITS PRECIPITATION

FT OF CALCIUM PHOSPHATE SALTS).  
 SQ SEQUENCE 42 AA; 5207 MW; E0630E5B184C506 CRC64;  
 Query Match 22.3%; Score 70.5; DB 1; Length 42;  
 Best Local Similarity 46.2%; Pred. No. 0.11;  
 Matches 18; Conservative 4; Mismatches 12; Indels 5; Gaps 2;

OY 20 DSSEKRRRRKKHRRGFGQYQYRPLNTPPAYPP 58  
 DB 1 DSSEKRRRRRRRDEG---RYGYPQ--FVPPPLYPQP 34

RESULT 7  
 ID\_CASK\_RABIT STANDARD; PRT; 180 AA.  
 AC P33618; Q9TV51;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Kappa casein precursor.  
 GN CSN3 OR CSN10 OR CSNK.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white; TISSUE=Mammary gland;  
 RX MEDLINE=94059363; PubMed=8240676;  
 RA Bosze Z., Devinyo E., Pulsant C., Fontaine M.L., Houdebine L.M.;  
 RT "Characterization of rabbit kappa-casein cDNA: control of kappa-casein  
 RT gene expression in vivo and in vitro.";  
 RL J. Mol. Endocrinol. 11:9-17(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white;  
 RX MEDLINE=97017124; PubMed=8863725;  
 RA Baranyi M., Aszodi A., Devinyo E., Fontaine M.L., Houdebine L.M.,  
 RA Bosze Z.;  
 RT "Structure of the rabbit kappa-casein encoding gene: expression of  
 RT the cloned gene in the mammary gland of transgenic mice.";  
 RL Gene 174:27-34(1996).  
 RN [3]  
 RP SEQUENCE OF 31-180 FROM N.A.  
 RC STRAIN=New Zealand white;  
 RA Hirtip L., Devinyo E., Rat P., Baranyi M., Fontaine M.L., Bosze Z.;  
 RT "Polymorphism of the rabbit kappa-casein gene. Presence or absence of  
 RT both 1800 nt and 100 nt in two intronic regions containing  
 RT microsatellites and similar to LINE.";  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 55-167 FROM N.A.  
 RC STRAIN=New Zealand white; TISSUE=Mammary gland;  
 RA Dawson S.P.;  
 RL Thesis (1993), Nottingham University, U.K.  
 CC -1- FUNCTION: Kappa casein stabilizes micelle formation, preventing  
 CC casein precipitation in milk.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.  
 CC -1- SIMILARITY: Belongs to the kappa-casein family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z18243; CA479145.1; -  
 DR EMBL; U44057; AAC48796.1; -  
 DR EMBL; U44055; AAC48796.1; JOINED.  
 DR EMBL; U44056; AAC48796.1; JOINED.

```

DR EMBL; AF033505; AAD48079.1; -.
DR EMBL; X76910; CA54232.1; -.
DR PIR; JC4955; JC4955.
DR InterPro; IPR00117; Casein_kappa.
DR Pfam; PF00997; Casein_Kappa; 1.
DR ProDom; PD003689; Casein_Kappa; 1.
DR Milk; Glycoprotein; Phosphorylation; Signal.
FT SIGNAL 1 21
FT CHAIN 22 180
FT SITE 116 117
FT SITE 116 117
FT CONFLICT 56 57
FT CONFLICT 166 166
SQ SEQUENCE 180 AA; 20342 MW; A1F058E01725F56E CRC64;

Query Match 21.0%; Score 66.5; DB 1; Length 180;
Best Local Similarity 31.7%; Pred. No. 1.2;
Matches 20; Conservative 10; Mismatches 24; Indels 9; Gaps 3;

Qy 1 MKTFFVFF-IMALTIAMIRADSESEKRRKKHGRGYFOOYQ-----QRYPLNYP 52
Db 2 MKHFLVNLVNLAVLPPLADIQNQEQTTGRENERLPHOYTATYIFVHYVNRYP-QYE 60

Qy 53 PAY 55
Db 61 PSY 63

RESULT 8
BRD4_HUMAN
ID BRD4_HUMAN STANDARD; PRT; 1362 AA.
AC 0608B5; Q96PD3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bromodomain-containing protein 4 (HUNK1 protein).
GN BRD4 OR HUNK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RA French C.A., Fletcher J.A.;
RT "Human BRD4 protein.";
RN Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RC [2]
RP TISSUE=Placenta;
RA Weber B.;
RN Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: Contains 2 bromodomains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF386649; AAL26987.1; -.
DR EMBL; Y12059; CA72780.1; -.
DR HSSP; Q92831; 1891.
DR Genem; HGNC:13575; BRD4.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
KM Bromodomain; Repeat; Nuclear protein.

```

```

FT DOMAIN 75 147
FT DOMAIN 368 440
FT DOMAIN 535 594
FT DOMAIN 692 717
FT DOMAIN 703 714
FT DOMAIN 738 743
FT DOMAIN 757 761
FT DOMAIN 764 770
FT DOMAIN 771 775
FT DOMAIN 776 783
FT DOMAIN 954 964
FT DOMAIN 974 986
FT DOMAIN 1011 1014
FT DOMAIN 1028 1033
FT DOMAIN 1283 1303
FT DOMAIN 1301 1308
FT DOMAIN 1335 1338
FT CONFLICT 720 721
SQ SEQUENCE 1362 AA; 152219 MW; D52EFELCP960907 CRC64;

Query Match 20.7%; Score 65.5; DB 1; Length 1362;
Best Local Similarity 39.0%; Pred. No. 11;
Matches 19; Conservative 5; Mismatches 13; Indels 11; Gaps 3;

Qy 19 ADSSEKRRH--RRKKHGRGYFOOYQ-----PYQRYPLNYPYPPYPP 58
Db 722 APKSKKKKGRGRRGQKKHHHHHQQMQADAPAPVQQRPP---PRQQRPP 766

RESULT 9
ID YD3D SCHPO STANDARD; PRT; 533 AA.
AC 01027;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical RNA-binding protein Cl3G7.13c in chromosome 1.
GN SPAC13G7.13c OR SPAC6C3.01c.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Baeham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben U., Gymnopoulos B.,
RA Welffens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabell C., Fuchs M., Fritzc C., Holzer E., Moesli D., Hilbert H.,
RA Borzys K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut R., Farnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
RA Gilbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Baga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Sipakovski G.V., Ussery D., Barrell B.G., Nutsche P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RN Nature 415:871-880(2002).

```





RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP SEQUENCE OF 25-182.  
 RX MEDLINE=65258132; PubMed=4018271;  
 RA Brignon G., Chetourou A., Ribadeau-Dumas B.;  
 RT "Preparation and amino acid sequence of human kappa-casein.";  
 RL FEBS Lett. 188:48-54(1985).  
 RN [6]  
 RP CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 118-182.  
 RX MEDLINE=8114144; PubMed=7460900;  
 RA Fat A.M., Jolles J., Aubert J.-P., Loucheux-Lefebvre M.H., Jolles P.;  
 RT "Localisation and importance of the sugar part of human casein.";  
 RL Eur. J. Biochem. 111:333-339(1980).  
 RN [7]  
 RP STRUCTURE BY NMR OF 108-125.  
 RX MEDLINE=99207197; PubMed=10191473;  
 RA Plowman J.E., Creamer L.K., Liddell M.J., Cross J.J.;  
 RT "Structural features of a peptide corresponding to human kappa-casein  
 RT residues 84-101 by 1H-nuclear magnetic resonance spectroscopy.";  
 RL J. Dairy Res. 66:53-63(1999).  
 CC -1- FUNCTION: Kappa casein stabilizes micelle formation, preventing  
 CC casein precipitation in milk.  
 CC -1- SUBUNIT: HETEROMULTIMERS COMPOSED OF ALPHA-S1 CASEIN AND KAPPA  
 CC CASEIN LINKED BY DISULFIDE BONDS.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.  
 CC -1- PTM: The N-terminus is blocked.  
 CC -1- SIMILARITY: Belongs to the kappa-casein family.  
 CC -1- DATABASE: NAME=Protein Spotlight;  
 CC NOTE=Issue 16 of November 2001;  
 CC WWW="http://www.expasy.org/spotlight/articles/spl1016.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC  
 DR EMBL; M73628; AAA59456.1; -;  
 DR EMBL; X66417; AAA47048.1; -;  
 DR EMBL; U51899; AAC50772.1; -;  
 DR EMBL; BC010935; AAH10935.1; -;  
 DR PIR; JC4957; KKHU.  
 DR GlycoSuiteDB; P07498; -;  
 DR Genew; HGNC:2446; CSN3.  
 DR MIM; 601695; -;  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0007582; P:physiological processes; NAS.  
 DR InterPro; IPR000117; Casein\_kappa.  
 DR Pfam; PF00997; Casein\_kappa; 1.  
 DR ProDom; PD003689; Casein\_kappa; 1.  
 DR KMW; Milk; Phosphorylation; Glycoprotein; Signal.  
 FT CHAIN 1 20  
 FT SIGNAL 1 20  
 FT SITE 21 182 KAPPA CASEIN.  
 FT CARBOHYD 117 118 CLEAVAGE (BY CHYMOSIN (=RENNIN)).  
 FT CARBOHYD 133 133 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 143 143 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 148 148 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 151 151 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 157 157 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 169 169 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 172 172 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 174 174 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 178 178 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 181 181 O-LINKED (GALNAC. . .).  
 FT CONFLICT 25 27 OKO -> EOK (IN REF. 5).  
 FT CONFLICT 75 75 Y -> C (IN REF. 4).  
 FT CONFLICT 110 110 R -> L (IN REF. 3).  
 FT CONFLICT 145 145 A -> T (IN REF. 1).  
 FT

FT CONFLICT 153 153 D -> N (IN REF. 5).  
 FT CONFLICT 162 162 S -> T (IN REF. 5 AND 6).  
 FT CONFLICT 173 173 T -> P (IN REF. 5 AND 6).  
 FT CONFLICT 178 181 TPPT -> PPTS (IN REF. 5 AND 6).  
 SQ SEQUENCE 182 AA; 20305 MW; 8A1609D2D5C94895 CRC64;  
 Query Match 19.3%; Score 63; DB 1; Length 182;  
 Best Local Similarity 28.8%; Pred. No. 3.1;  
 Matches 17; Conservative 13; Mismatches 25; Indels 4; Gaps 3;  
 QY 1 MKLFIYFV-IMAILMIRADSEKRRKRRKHHGYPQYQYQYPLNY-PNAVPF 57  
 DB 1 MKSFLVWVALTLPLAVEVONKOPACHENDEREFYKTAAY--VPMYVYVNSY 57  
 RESULT 12  
 CC37\_TERFL  
 ID CC37\_TERFL STANDARD; PRT; 377 AA.  
 AC Q9DQ07;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting  
 DE subunit) (p50Cdc37).  
 GN CDC37.  
 OS Tetraodon fluviatilis (Puffer fish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontidae; Tetraodon.  
 ON NCBI\_TaxID=47145;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20525449; PubMed=11072077;  
 RA Tsai S.C., Leu J.H., Chou C.M., Yeh M.S., Huang F.L., Huang C.J.;  
 RT "Genomic organization and the promoter region of the round-spotted  
 RT pufferfish (Tetraodon fluviatilis) CDC37 gene.";  
 RL Biochim. Biophys. Acta 1494:144-148(2000).  
 CC -1- FUNCTION: With Hsp90 it forms a complex that binds to several  
 CC kinases, resulting in stabilization and promotion of their  
 CC activity (By similarity).  
 CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with CDK4 and CDK6  
 CC (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE CDC37 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC  
 DR EMBL; AF091237; AAC00066.1; -;  
 DR InterPro; IPR004918; Cdc37.  
 DR Pfam; PF03234; Cdc37; 1.  
 DR KMW; Chapterone.  
 SQ SEQUENCE 377 AA; 44482 MW; F8BD12F81F786A6D CRC64;  
 Query Match 19.3%; Score 61; DB 1; Length 377;  
 Best Local Similarity 29.2%; Pred. No. 10;  
 Matches 14; Conservative 7; Mismatches 15; Indels 12; Gaps 1;  
 QY 17 IRADSEBKRRKRRKRRK-----HKGFGQYQYQYQYPLNY 52  
 DB 135 VKADSADTEEEKKRGKTFVERYERKQIKHFGMLHRWDSQRYLSDNP 182  
 RESULT 13  
 CASB\_HUMAN  
 ID CASB\_HUMAN STANDARD; PRT; 226 AA.  
 AC P05814;

DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Beta casein precursor.  
GN CSN2 OR CASB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast;  
RA Menon R.S.;  
RL Submitted (OCT-1989) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=90353560; PubMed=2387396;  
RA Loeferer B., Bergstrom S., Andersson Y., Hjalmarsson K.,  
RA Sundqvist A.K., Hernell O.;  
RT "Cloning and sequencing of a cDNA encoding human milk beta-casein.";  
RL FEBS Lett. 269:153-156(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA MEDLINE=94156198; PubMed=8112603;  
RA Hansson L., Edlund A., Johansson T., Hernell O., Stromqvist M.,  
RA Lindqvist S., Loeferer B., Bergstrom S.;  
RT "Structure of the human beta-casein encoding gene.";  
RL Gene 139:193-199(1994).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Kwiatkowski D.J.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 161-226 FROM N.A.  
RC TISSUE=Breast;  
RA MEDLINE=89240053; PubMed=2717418;  
RA Menon R.S., Ham R.G.;  
RT "Human beta-casein: partial cDNA sequence and apparent polymorphism.";  
RL Nucleic Acids Res. 17:2869-2869(1989).  
RN [6]  
RP SEQUENCE OF 16-226.  
RA MEDLINE=84185624; PubMed=6715339;  
RA Greenberg R., Groves M.L., Dower H.J.;  
RT "Human beta-casein. Amino acid sequence and identification of  
RT phosphorylation sites";  
RL J. Biol. Chem. 259:5132-5138(1984).  
CC -1- FUNCTION: IMPORTANT ROLE IN DETERMINATION OF THE SURFACE  
CC PROPERTIES OF THE CASEIN MICELLES.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.  
CC -1- SIMILARITY: BELONGS TO THE BETA-CASEIN FAMILY.  
CC -1- DATABASE: NAMB=Protein SpotLight;  
CC NOTE=Issue 16 of November 2001;  
CC WWW="http://www.expasy.org/spotlight/articles/sp1016.html".  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; X17070; CAA4916.1; -  
CC EMBL; X13766; CAA32017.1; -  
CC EMBL; AF027807; AAC82978.1; -  
CC EMBL; X55739; CAA39270.1; -  
CC EMBL; A24287; CAA01728.1; -  
CC EMBL; A30262; CAA02017.1; -  
CC PIR; I53730; KBHU.  
CC  
CC Gene; HGNC:2447; CSN2.

DR MIM; 115460; -  
DR GO; GO:0005509; F:calcium ion binding activity; TAS.  
DR GO; GO:0004857; F:enzyme inhibitor activity; TAS.  
DR GO; GO:0005189; F:milk protein; TAS.  
DR GO; GO:0006816; F:calcium ion transport; TAS.  
DR InterPro; IPR001588; Casein.  
DR Pfam; PF00363; casein; 1.  
DR PROSITE; PS00306; CASEIN ALPHA BETA; 1.  
KW Milk; phosphorylation; glycoprotein; signal.  
FT SIGNAL 1 15  
FT CHAIN 16 226  
FT MOD\_RES 18 18 BETA CASEIN.  
FT MOD\_RES 21 21 PHOSPHORYLATION.  
FT MOD\_RES 23 23 PHOSPHORYLATION.  
FT MOD\_RES 24 24 PHOSPHORYLATION.  
FT MOD\_RES 25 25 PHOSPHORYLATION.  
FT CONFLICT 30 30 T -> P (IN REF. 6).  
FT CONFLICT 34 34 MISSING (IN REF. 2).  
FT CONFLICT 48 50 EDE -> TDO (IN REF. 6).  
FT CONFLICT 120 120 S -> Q (IN REF. 6).  
FT CONFLICT 133 133 L -> V (IN REF. 1).  
FT CONFLICT 140 140 H -> Q (IN REF. 1).  
FT CONFLICT 149 149 L -> S (IN REF. 6).  
FT CONFLICT 173 173 Q -> E (IN REF. 6).  
FT CONFLICT 182 184 QVV -> EVL (IN REF. 6).  
FT CONFLICT 188 188 O -> V (IN REF. 6).  
FT CONFLICT 207 207 T -> P (IN REF. 6).  
FT CONFLICT 214 222 TOPLAVERN -> PEPSTTZAHH (IN REF. 6).  
SQ SEQUENCE 226 AA; 25382 MW; 2619C524EA1358E8 CRC64;  
  
Query Match 19.1%; Score 60.5; DA 1; Length 226;  
Best Local Similarity 28.9%; Pred. No. 7.2;  
Matches 22; Conservative 9; Mismatches 24; Indels 21; Gaps 3;  
  
QY 1 MKLIFVFIMALLMIRADS-----SEKRRKKKRGVFOQ-----YQ 42  
DB 1 MKVLIIACVALLARETIESLSSESESTYEKQVEKVED---QQGDEHDKIYP 57  
QY 43 PYQRYPLNYPVPPPP 58  
DB 58 SFQPPPLIYFVEVP 73  
  
RESULT 14  
C1KA\_BACTM  
ID C1KA\_BACTM STANDARD; PRT; 1215 AA.  
AC Q45715;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pesticidal crystal protein cryIa (insecticidal delta-endotoxin  
DE cryIa(a)) (crystalline entomocidal protoxin) (137 kDa crystal protein).  
GN CRYIKA OR CRYIK(A) OR CRYIK.  
OS Bacillus thuringiensis (subsp. morrisoni).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1441;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=F190;  
RX MEDLINE=96102856; PubMed=8586263;  
RA Koo B.T., Park S.-H., Choi S.-K., Shin B.-S., Kim J.I., Yu J.H.;  
RT "Cloning of a novel crystal protein gene cryIa from Bacillus  
RT thuringiensis subsp. morrisoni";  
RL FEMS Microbiol. Lett. 134:159-164(1995).  
CC -1- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT  
CC EPITHELIAL CELLS OF INSECTS. SELECTIVELY TOXIC TO ARTOGEBIA RAPAE  
CC AND NOT ACTIVE ON PUTELELLA XYLOSTELLA.  
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING  
CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART  
CC OF THE SPORE COAT.  
CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE  
CC N-TERMINUS.  
CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 13, 2003, 03:00:34 ; Search time 52 Seconds  
(without alignments)  
287.828 Million cell updates/sec

Title: US-10-079-754A-10

Perfect score: 316  
Sequence: 1 MKIFIFVFMALIMLRAD.....QQQPYQRYPLNYPAYPPF 58

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123.5	39.1	59	6	Q8HY86
2	75	23.7	192	5	Q9VKM5
3	72.5	22.9	336	5	002586
4	72	22.8	660	5	Q8MPZ9
5	68.5	21.7	336	5	Q95NFI
6	68	21.5	165	5	Q8N026
7	68	21.5	165	5	Q8N024
8	68	21.5	165	5	Q8N031
9	67.5	21.4	239	5	Q9WZRS
10	66	20.9	310	10	Q41210
11	66	20.9	386	3	014329
12	66	20.9	518	5	Q8T011
13	66	20.9	733	10	Q9LFR6
14	66	20.9	741	5	Q8T9G4
15	66	20.9	2252	5	Q9VZX4
16	65.5	20.7	447	16	Q9PHJ1

17	65.5	20.7	1203	5	Q45916
18	65	20.6	260	10	Q40053
19	65	20.6	282	5	Q9XWV4
20	65	20.6	321	5	Q17504
21	65	20.6	347	10	Q40055
22	65	20.6	1572	10	Q8H4N1
23	64	20.3	400	5	Q22029
24	64	20.3	2016	5	Q9VG78
25	63.5	20.1	75	5	Q8T3D9
26	63.5	20.1	177	16	Q8Z099
27	63.5	20.1	1124	10	Q8H1U2
28	63.5	20.1	1439	5	Q9NDN7
29	63	19.9	98	10	Q8S8M0
30	63	19.9	684	10	Q8H1O5
31	63	19.9	757	2	Q8GRA2
32	62.5	19.8	211	10	Q9LGS8
33	62.5	19.8	280	10	Q9FUW7
34	62.5	19.8	742	2	Q9ZB47
35	62.5	19.8	8094	5	Q8ILB9
36	62	19.6	153	10	Q9L1I5
37	62	19.6	166	5	Q8N021
38	62	19.6	699	5	Q8H1O9
39	62	19.6	719	5	Q25826
40	62	19.6	1534	4	Q9ULD9
41	62	19.6	1572	10	Q94DL2
42	61.5	19.5	233	6	Q9GKX3
43	61.5	19.5	294	5	Q9VXJ3
44	61.5	19.5	480	5	Q8T829
45	61.5	19.5	506	5	Q961E9

## ALIGNMENTS

### RESULT 1

ID	Q8HY86	PRELIMINARY;	PRT;	59 AA.
AC	Q8HY86;			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	Statherin.			
GN	STATH.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
OX	(1)			
RP	SEQUENCE FROM N. A.			
RC	TISSUE=Salivary gland;			
RA	Rijnkeels M., Elmltski L., Miller W., Rosen J.M.;			
RT	"Multi-species comparative sequence analysis of the casein gene			
RT	cluster region: A gene domain encoding epithelial secretory			
RT	proteins."			
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY154893; AA05581.1;			
DR	SEQUENCE 59 AA; 7093 MW; EE9866EBE0B21F25 CRC64;			

Query Match 39.1%; Score 123.5; DB 6; Length 59;  
Best Local Similarity 54.2%; Pred. No. 4.4e-08;  
Matches 32; Conservative 5; Mismatches 19; Indels 3; Gaps 3;

QY 1 MKIFIFVFMALIMLRADSSSEKRRKRRKRGYQYQY-QRYPLNYPAYPPF 58  
DB 1 MKIFFAFVFMALIMLRADSSSEKRRKRRKRGYQYQY-QRYPLNYPAYPPF 57

### RESULT 2

ID Q9VKM5 PRELIMINARY; PRT; 192 AA.  
AC Q9VKM5;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)  
 DE CG16743 protein (RH23514p).  
 GN CG16743  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton R.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,  
 RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasner K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Houston D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Koike C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Schejter F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Stykstra R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhang F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez C., Guarin H., Krommiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunco J., Paclob J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE003630; AAF53039.1; -;  
 DR EMBL: AY119164; AAF51024.1; -;  
 DR FlyBase: FBgn0032322; CG16743.  
 DR InterPro: IPR002965; P\_Rich\_extensn.  
 DR PRINTS: PRO1217; PRICHXTNSN.  
 SQ SEQUENCE 192 AA; 20751 MW; DD2363009D7DDEB CRC64;

Db 10 LTVLLQYEVFRDSDSSSSSGEKHKHSKTEHKYNYPAYPNPGYPPYOEMGCPYS 69  
 QY 41 --YQYQRYPLNPF-----PAYPF 57  
 Db 70 YLYNPWPYP-PYPQMGYPSPY 92  
 RESULT 3  
 ID 002586 PRELIMINARY; PRT; 336 AA.  
 AC 002586; 002585;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE Cysteine proteinase.  
 OS Spilometra erinaceaeuropaei.  
 OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;  
 OC Platyhelminthes; Diphyllobothriidae; Spilometra.  
 OX NCBI\_TaxID=99802;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97077424; PubMed=8919991;  
 RA Liu D.W., Kato H., Nakamura T., Sugane K.;  
 RT "Molecular cloning and expression of the gene encoding a cysteine  
 RT proteinase of Spilometra erinaceae";  
 RL Mol. Biochem. Parasitol. 76:11-21(1996).  
 DR EMBL: D63670; BA09821.1; -;  
 DR EMBL: D63670; BA09820.1; -;  
 DR HSP; P07711; ICDL.  
 DR InterPro: IPR000668; peptidase\_C1.  
 DR InterPro: IPR000169; SHPOT\_acstle.  
 DR Pfam: PF00112; Peptidase\_C1; 1.  
 DR PRINTS: PRO0705; PAPAIN.  
 DR ProDom: PD000158; Peptidase\_C1; 1.  
 DR SMART: SM00645; Pept\_C1; 1.  
 DR PROSITE: PS00640; THIOL\_PROTEASE\_ASN; 1.  
 DR PROSITE: PS00139; THIOL\_PROTEASE\_CYS; 1.  
 DR PROSITE: PS00639; THIOL\_PROTEASE\_HIS; 1.  
 KW Hydrolase; Protease; Thiol protease.  
 SQ SEQUENCE 336 AA; 37871 MW; 7072283F52C0FDB CRC64;  
 Query Match 22.9%; Score 72.5; DB 5; Length 336;  
 Best Local Similarity 26.0%; Pred. No. 0.57;  
 Matches 20; Conservative 10; Mismatches 14; Indels 33; Gaps 2;  
 QY 1 MKLFIYFVFMALILAMIRAD-----SSEKRRKRRK----- 31  
 Db 1 MKFVIYVAFLLTLTCRGSTSEYVRLRMKMLAFKKEYSSSEELHRRRAFFNNL 60  
 QY 32 ----KHHRGYFOOQY 44  
 Db 61 DFTIRHQRYYQOLEST 77  
 RESULT 4  
 ID 08MPZ9 PRELIMINARY; PRT; 660 AA.  
 AC 08MPZ9;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE R07B5.9 protein.  
 GN R07B5.9.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kelly P.F.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=9069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology."  
RL Science 282:2012-2018(1998).  
DR EMBL; Z72512; CAD4413.1; -.  
DR WormPep; R07B5.9; CEJ1575.  
SQ SEQUENCE 660 AA; 72956 MW; 8DCFB1E7FEC0B8D3 CRC64;

Query Match  
Best Local Similarity 22.8%; Score 72; DB 5; Length 660;  
Matches 17; Conservative 10; Mismatches 13; Indels 22; Gaps 2;

QY 18 RADSEKKRRKRRKHHRGYPQOYQRYOR-----YPLN-----YPPAY 55  
DB 541 RSESATQRTKARQOHQOHOQOQPOPOQRIAPGVPGVHPOMQPMNMMMPAYPPY 600  
QY 56 PF 57  
DB 601 PY 602

RESULT 5  
ID Q95NF1 PRELIMINARY; PRT; 336 AA.  
AC Q95NF1;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Pterocercoid growth factor-2/cysteine protease (Pterocercoid growth  
factor/cysteine protease).  
GN PGF-E2 OR PGF-E.  
OS Spirometra erinaceieuropaei.  
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;  
OC Pseudophyllidea; Diphyllidobothriidae; Spirometra.  
OX NCBI\_TaxId=99802;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hirai K., Wang H., Fukumoto S., Sato K., Tadamoto S.;  
RT "An unusual mRNA for pterocercoid growth factor (PGF-E2)/cysteine  
protease with UUG initiation code from pterocercoids of Spirometra  
erinaceieuropaei."  
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Hirai K., Wang H., Fukumoto S., Sato K., Tadamoto S.;  
RT "Molecular cloning and characterization of a growth factor from the  
RT pterocercoids of Spirometra erinaceieuropaei."  
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB055646; BAB62718.1; -.  
DR EMBL; AB055411; BAB62718.1; -.  
DR InterPro; IPR000668; Peptidase\_C1.  
DR InterPro; IPR000169; SHprot\_acidite.  
DR Pfam; PF00112; Peptidase\_C1.1.  
DR PRINTS; PR00705; PAPAIN\_1.  
DR ProDom; PD000158; Peptidase\_C1.1.  
DR SMART; SM00645; Pept\_C1.1.  
DR PROSITE; PS00640; THIOL\_PROTEASE\_ASN; 1.  
DR PROSITE; PS00139; THIOL\_PROTEASE\_CYS; 1.  
DR PROSITE; PS00639; THIOL\_PROTEASE\_HIS; 1.  
KW Hydrolase; Protease; Thiol protease.  
SQ SEQUENCE 336 AA; 37775 MW; FD70B02F1F12333 CRC64;

Query Match  
Best Local Similarity 21.7%; Score 68.5; DB 5; Length 336;  
Matches 19; Conservative 10; Mismatches 15; Indels 33; Gaps 2;

QY 1 WKIFVFVFMALILMIRAD-----SSEKRRK----- 31  
DB 1 MKEFVIVAFLEFLLTVCCSTGSEYVRRELKMKWAKLAFKEYFSSSEELHKKRAFFNNL 60  
QY 32 ----KHHRGYPQOYQRY 44

DB 61 DFIIRHNRQYVQQLSESY 77

RESULT 6  
ID Q8N026 PRELIMINARY; PRT; 165 AA.  
AC Q8N026;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hunchback (Fragment).  
GN HB.  
OS Drosophila quadraria.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=67532;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schawaroch V.A.;  
RT "Phylogeny of a paradigm lineage: the Drosophila melanogaster species  
RT group (Diptera: Drosophilidae).";  
RL Biol. J. Linn. Soc. 76:21-37(2002).  
DR EMBL; AF461343; AAM28799.1; -.  
DR FlyBase; FBgn0062735; Dqua\hb.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 165 AA; 18261 MW; DFBE744F1795604A CRC64;

Query Match  
Best Local Similarity 21.5%; Score 68; DB 5; Length 165;  
Matches 12; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

QY 19 ADSEKKRRKRRKHHRGYPQOYQRYQRYPLN 50  
DB 16 AASQLEQFLKQOQHQQOHHQOHHQOHHQOHPMD 47

RESULT 7  
ID Q8N024 PRELIMINARY; PRT; 165 AA.  
AC Q8N024;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hunchback (Fragment).  
GN HB.  
OS Drosophila triauraria.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=67534;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schawaroch V.A.;  
RT "Phylogeny of a paradigm lineage: the Drosophila melanogaster species  
RT group (Diptera: Drosophilidae).";  
RL Biol. J. Linn. Soc. 76:21-37(2002).  
DR EMBL; AF461345; AAM28801.1; -.  
DR FlyBase; FBgn0062721; Dcra\hb.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 165 AA; 18261 MW; DFBE744F1795604A CRC64;

Query Match  
Best Local Similarity 21.5%; Score 68; DB 5; Length 165;  
Matches 12; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

QY 19 ADSEKKRRKRRKHHRGYPQOYQRYQRYPLN 50  
DB 16 AASQLEQFLKQOQHQQOHHQOHHQOHHQOHPMD 47





DT 01-JAN-1999 (TrEMBLrel. 09, Created)  
DT 01-JAN-1999 (TrEMBLrel. 09, last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
DE Zinc-regulated protein.  
GN SPBC169.14C.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
NCBI\_Taxid=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972h-;  
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z93759; CAB16906.1; -.  
DR GeneDB; SPombe; SPBC169.14C; -.  
DR InterPro; IPR002524; Cation\_efflux.  
DR Pfam; PF01545; Cation\_efflux; 1.  
SQ SEQUENCE 386 AA; 43089 MW; BCACA29DC57A8CD7 CRC64;

Query Match 20.9%; Score 66; DB 3; Length 386;  
Best Local Similarity 36.6%; Pred. No. 4.2;  
Matches 15; Conservative 4; Mismatches 18; Indels 4; Gaps 1;

QY 22 SEKKRRKKRHHRGYFOOYQ---PYQRYPLNYPAYPP 58  
DB 41 SVDTPTKHKHGHKHSLSHGYFLPKNQPLEIPASYIP 81

## RESULT 12

Q8T011 PRELIMINARY; PRT; 518 AA.  
ID Q8T011  
AC Q8T011  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)  
DE HL03404D.  
GN CG2083.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RA Stapleton M., Brokstein P., Hong J., Agbayani A., Carlson J.,  
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celnik S.,  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; A0069301; AAL39446.1; -.  
DR FlyBase; FBgn0035376; CG2083.  
SQ SEQUENCE 518 AA; 55827 MW; 91A20B166CB289 CRC64;

Query Match 20.9%; Score 66; DB 5; Length 518;  
Best Local Similarity 46.2%; Pred. No. 5.6;  
Matches 12; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 33 HHRGYFOOYQRYPLNYPAYPP 58  
DB 67 YNYGYHOAYSPLYGMYPQOTPPPP 92

## RESULT 13

Q9LFR6 PRELIMINARY; PRT; 733 AA.  
ID Q9LFR6  
AC Q9LFR6  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)

DE Hypothetical 83.2 kDa protein.  
GN P2G14.10.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
NCBI\_Taxid=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,  
RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL391146; CAC01808.1; -.  
DR InterPro; IPR001258; NHL.  
DR InterPro; IPR001412; tRNA-synt\_1.  
DR Pfam; PF01436; NHL; 2.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 733 AA; 83158 MW; B71F5057480D5E6 CRC64;

Query Match 20.9%; Score 66; DB 10; Length 733;  
Best Local Similarity 32.6%; Pred. No. 7.9;  
Matches 14; Conservative 8; Mismatches 21; Indels 0; Gaps 0;

QY 6 FVFIMALLIAMRADSSSEKKRHHRGYFOOYQRYPP 48  
DB 240 FGMLALIVRRVSLFSSSSHDTKSRHVAPSMATAPORYP 282

## RESULT 14

Q8T9G4 PRELIMINARY; PRT; 741 AA.  
ID Q8T9G4  
AC Q8T9G4  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)  
DE SD0233BP.  
GN CG2083.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stapleton M., Brokstein P., Hong J., Agbayani A., Carlson J.,  
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celnik S.,  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; A0069780; AAL39925.1; -.  
DR FlyBase; FBgn0035376; CG2083.  
SQ SEQUENCE 741 AA; 78400 MW; 1F9A808A990C49DC CRC64;

Query Match 20.9%; Score 66; DB 5; Length 741;  
Best Local Similarity 46.2%; Pred. No. 8;  
Matches 12; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 33 HHRGYFOOYQRYPLNYPAYPP 58  
DB 293 YNYGYHOAYSPLYGMYPQOTPPPP 318

## RESULT 15

Q9VZX4 PRELIMINARY; PRT; 2252 AA.  
ID Q9VZX4  
AC Q9VZX4  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE CG2083 protein.  
 GN CG2083.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC C1 TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brodtier P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
 RA Fowler C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jaisi M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Lin X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradscky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE003476; AAF47691.2; -  
 DR FlyBase: FBgn0035376; CG2083.  
 DR InterPro: IPR007087; Znf.C2H2.  
 DR PROSITE: PS00028; ZINC.FINGER\_C2H2\_1; 1.  
 SQ SEQUENCE 2252 AA; 24387 MW; C2DCBF03EA70913C CRC64;  
 Query Match 20.9%; Score 66; DB 5; Length 2252;  
 Best local Similarity 46.2%; Pred. No. 24;  
 Matches 12; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
 QY 33 HHRGYFQGYQYQYRYPPLNPPAYFFP 58  
 DB 1801 YNNGYHQAYSPYGMVFPQGTPTPTPTPP 1826  
 Search completed: December 13, 2003, 03:08:41  
 Job time : 55 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 13, 2003, 03:09:15 / Search time 1493 Seconds  
(without alignments)  
944.179 Million cell updates/sec

Title: US-10-079-754A-10  
Perfect score: 316  
Sequence: 1 MKIFIFVFMALILAMIRAD.....QQYQYQRYPLNYPAPYFP 58

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 segs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO/US10079754/runat\_12122003\_145752\_13028/app\_query.fasta\_1.199  
-DB=EST -QFMT=fastad -SUFFIX=1stc -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCMALIGN=200 -THR SCORE=Pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10079754@cgn2\_1\_12810@runat\_12122003\_145752\_13028 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEOQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_tod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1.\*

29: gb\_gss2.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	36.1	569	14 CB959043	CB959043 AGENCOURT
2	113	35.8	845	14 CB958207	CB958207 AGENCOURT
3	111	35.1	779	14 CB987499	CB987499 AGENCOURT
4	110	34.8	286	9 AA376683	AA376683 EST89136
5	110	34.8	304	9 AA376720	AA376720 EST89178
6	110	34.8	326	14 CB958892	CB958892 AGENCOURT
7	110	34.8	331	14 CB955620	CB955620 AGENCOURT
8	110	34.8	341	9 AA376695	AA376695 EST89151
9	110	34.8	357	14 CB955686	CB955686 AGENCOURT
10	110	34.8	426	9 AA376680	AA376680 EST89133
11	110	34.8	441	14 CB957572	CB957572 AGENCOURT
12	110	34.8	457	9 AA376682	AA376682 EST89135
13	110	34.8	496	2 HSM075323	Bx485136 Homo sapi
14	110	34.8	506	2 HSM075180	Bx484993 Homo sapi
15	110	34.8	517	10 BG188342	BG188342 RST7360 A
16	110	34.8	517	14 CB986649	CB986649 AGENCOURT
17	110	34.8	519	10 AW951480	AW951480 EST86350
18	110	34.8	543	14 CB958742	CB958742 AGENCOURT
19	110	34.8	550	14 CB987677	CB987677 AGENCOURT
20	110	34.8	553	14 CB985810	CB985810 AGENCOURT
21	110	34.8	554	14 CB985279	CB985279 AGENCOURT
22	110	34.8	558	14 CB956225	CB956225 AGENCOURT
23	110	34.8	558	14 CB985451	CB985451 AGENCOURT
24	110	34.8	559	14 CB955807	CB955807 AGENCOURT
25	110	34.8	559	14 CB956351	CB956351 AGENCOURT
26	110	34.8	559	14 CB957201	CB957201 AGENCOURT
27	110	34.8	559	14 CB958349	CB958349 AGENCOURT
28	110	34.8	559	14 CB958931	CB958931 AGENCOURT
29	110	34.8	559	14 CB958367	CB958367 AGENCOURT
30	110	34.8	560	14 CB958121	CB958121 AGENCOURT
31	110	34.8	560	14 CB956175	CB956175 AGENCOURT
32	110	34.8	561	14 CB955583	CB955583 AGENCOURT
33	110	34.8	561	14 CB956887	CB956887 AGENCOURT
34	110	34.8	561	14 CB957236	CB957236 AGENCOURT
35	110	34.8	561	14 CB957288	CB957288 AGENCOURT
36	110	34.8	561	14 CB957497	CB957497 AGENCOURT
37	110	34.8	561	14 CB985128	CB985128 AGENCOURT
38	110	34.8	562	14 CB955599	CB955599 AGENCOURT
39	110	34.8	562	14 CB956057	CB956057 AGENCOURT
40	110	34.8	562	14 CB956247	CB956247 AGENCOURT
41	110	34.8	562	14 CB956265	CB956265 AGENCOURT
42	110	34.8	562	14 CB959444	CB959444 AGENCOURT
43	110	34.8	562	14 CB9844703	CB9844703 AGENCOURT
44	110	34.8	562	14 CB987570	CB987570 AGENCOURT
45	110	34.8	563	14 CB955684	CB955684 AGENCOURT

ALIGNMENTS

RESULT 1  
CB959043  
LOCUS  
DEFINITION CB959043 569 bp mRNA linear EST 29-APR-2003  
AGENCOURT 13665470 NIH MGC 184 Homo sapiens cDNA clone  
IMAGE:30353949 5', mRNA sequence.  
ACCESSION  
CB959043  
VERSION  
CB959043.1 GI:30215159  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 569)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: NDCM151 row: h column: 22  
 High quality sequence stop: 553.  
 Location/Qualifiers

FEATURES  
 source 1..569  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30353949"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIH MGC 184"  
 /note="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site\_1: SfiI (ggccatcaggcc); Site\_2: SfiI (ggccgcctcggcc); Library is oligo-dT primed and directionally cloned. CDNA was prepared from a glandular pool of tissues from thyroid, parathyroid, adrenal, cortex and pineal gland. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATCTAGAGCGCGGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.38 kb (range 0.60-3.5 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 178 a 121 c 81 g 188 t 1 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 0.00943 Length: 569  
 Score: 114.00 Matches: 28  
 Percent Similarity: 63.64% Conservative: 7  
 Best Local Similarity: 50.91% Mismatches: 18  
 Query Match: 36.08% Indels: 2  
 Gaps: 2

US-10-079-754a-10 (1-58) x CB959043 (1-569)

QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaASP 20  
 Db 79 ATGAAGTTCTTGTGCTTCATCTTGCGCTTCATGTTCCATGATGAGCGTAT 138  
 QY 21 SerSerGluGluLysArgHisArgLysArgLysHisHisArgGlyTyrPheGlnGln 40  
 Db 139 TCATCTGAAGAAGAAATTTTGCCTAGAAATGGAAGATTCGGTTATGGGTAT--GGCCCT 195  
 QY 41 TyGlnProTyrGlnArgTyrProLeuAsnTyrProProAlaTyr 55  
 Db 196 TATCAGCAGTTCAGAACACACCTA--TACCCACACCATAC 237

RESULT 2  
 CB958207 845 bp mRNA linear EST 29-APR-2003  
 LOCUS AGENCOURT 13800675 NIH\_MGC\_184 Homo sapiens CDNA clone  
 DEFINITION IMAGE:30352463 5', mRNA sequence.  
 ACCESSION CB958207  
 VERSION CB958207.1 GI:30214323  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 845)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: NDCM151 row: j column: 24  
 High quality sequence stop: 211.  
 Location/Qualifiers

FEATURES  
 source 1..845  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30352463"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIH MGC 184"  
 /note="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site\_1: SfiI (ggccatcaggcc); Site\_2: SfiI (ggccgcctcggcc); Library is oligo-dT primed and directionally cloned. CDNA was prepared from a glandular pool of tissues from thyroid, parathyroid, adrenal, cortex and pineal gland. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATCTAGAGCGCGGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.38 kb (range 0.60-3.5 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 246 a 193 c 153 g 251 t 2 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 0.0164 Length: 845  
 Score: 113.00 Matches: 28  
 Percent Similarity: 60.34% Conservative: 7  
 Best Local Similarity: 48.28% Mismatches: 19  
 Query Match: 35.76% Indels: 4  
 Gaps: 2

US-10-079-754a-10 (1-58) x CB958207 (1-845)

QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaASP 20  
 Db 76 ATGAAGTTCTTGTGCTTCATCTTGCGCTTCATGTTCCATGATGAGCGTAT 135  
 QY 21 SerSerGluGluLysArgHisArgLysArgLysHisHisArgGlyTyrPheGlnGln 40  
 Db 136 TCATCTGAAGAAGAAATTTTGCCTAGAAATGGAAGATTCGGTTATGGGTAT--GGCCCT 192  
 QY 41 TyGlnPro-----TyrGlnArgTyrProLeuAsnTyrProProAlaTyr 55  
 Db 193 TATCAGCAGTTCAGAACACACCTATACCTACCTACCATCACCACACATAC 246

RESULT 3  
 CB987499 779 bp mRNA linear EST 01-MAY-2003  
 LOCUS AGENCOURT 13565248 NIH\_MGC\_184 Homo sapiens CDNA clone  
 DEFINITION IMAGE:30325523 5', mRNA sequence.  
 ACCESSION CB987499  
 VERSION CB987499.1 GI:30282019  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 779)



ACCESSION AA376720  
 VERSION AA376720.1 GI:2029038  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 304)  
 Adame,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gockayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Colton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geophagen,N.S., Glodok,A., Gheun,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmore,S.M., Merrick,J.M., Moreno-Palacios,R.F., McDonald,L.A., Nguyen,D.T., Pellingrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utecherback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dmke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Korak,D.L., Kunsch,C., Hungjun,J., Li,H., Weisner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M., and Venter,J.C.  
 Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
 Nature 377 (6547 Suppl), 3-174 (1995)  
 96026280  
 7566098  
 Other ESTs: EST89177 THC77891  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..304  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="ATCC (Inhost):181127"  
 /db\_xref="taxon:9606"  
 /dev\_stage="adult"  
 /clone\_lib="salivary gland"  
 /note="Organ: salivary gland; Vector: pBlueScript SK-;  
 Site\_1: EcorI; Site\_2: XhoI"  
 Site\_72 c 50 g 93 t  
 BASE COUNT 89 a 72 c  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.0159 Length: 304  
 Score: 110.00 Matches: 27  
 Percent Similarity: 61.82% Conservative: 7  
 Best Local Similarity: 49.09% Mismatches: 19  
 Query Match: 34.81% Indels: 2  
 Gaps: 2  
 DB: 9  
 US-10-079-754A-10 (1-58) x AA376720 (1-304)  
 Qy 1 MetLysIlePheValPheValPheIleMeValLeuIleLeuAlaMetIleArgAlaASP 20  
 Db 33 ATGAAGTTCCTTCTTTCCTTCATCTGCTCTCATGCTTCATGATGAGTGGCTGAT 92  
 Qy 21 SerSerGluGluGlyArgHisArgLysArgLysHisHisArgGlyTyrPheGlnGln 40  
 Db 93 TCATCTGAAGAAATTTTGGCTGAGAAATGGAAGATTCCGTTATGGTAT---GGCCCT 149

Qy 41 TyrGlnProTyrGlnArgTyrProLeuAenTyrProProAlaTyr 55  
 Db 150 TATCAGCCAGTTCACAGACACCACTA---TACCACACCACTAC 191  
 RESULT 6  
 CB958892  
 LOCUS  
 DEFINITION AGENCOURT 13784766 NIH MGC 184 Homo sapiens cDNA clone IMAGE:30352370 5', mRNA sequence.  
 ACCESSION CB958892  
 VERSION CB958892.1 GI:30215008  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 326)  
 NIH-MGC <http://imgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Straubeberg, Ph.D.  
 Email: c9abds-remail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Plate: NDCM151 row: 9 column: 03  
 High quality sequence stop: 310.  
 Location/Qualifiers  
 1..326  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30352370"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIH MGC 184"  
 /note="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site\_1: SfiI (ggccatctagc); Site\_2: SfiI (ggccgcctggcc); Library is oligo-dT primed and directionally cloned. cDNA was prepared from a glandular pool of tissues from thyroid, parathyroid, adrenal, cortex and pineal gland. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGAGCGCGCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.38 kb (range 0.60-3.5 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."  
 BASE COUNT 93 a 79 c 55 g 98 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.0168 Length: 326  
 Score: 110.00 Matches: 27  
 Percent Similarity: 61.82% Conservative: 7  
 Best Local Similarity: 49.09% Mismatches: 19  
 Query Match: 34.81% Indels: 2  
 Gaps: 2  
 DB: 14  
 US-10-079-754A-10 (1-58) x CB958892 (1-326)  
 Qy 1 MetLysIlePheValPheValPheIleMeValLeuIleLeuAlaMetIleArgAlaASP 20  
 Db 79 ATGAAGTTCCTTCTTTCCTTCATCTGCTCTCATGCTTCATGATGAGTGGCTGAT 138  
 Qy 21 SerSerGluGluGlyArgHisArgLysArgLysHisHisArgGlyTyrPheGlnGln 40  
 Db 139 TCATCTGAAGAAATTTTGGCTGAGAAATGGAAGATTCCGTTATGGTAT---GGCCCT 195

Qy 41 TygInProTyGInArgTyProLeuAsnTyProProAlaTyTyr 55  
 Db 196 TATCGAGCCAGTTCAGAACACACACTA---TACCACACACATAC 237

RESULT 7  
 LOCUS CB955620 331 bp mRNA linear EST 29-APR-2003  
 DEFINITION AGENCOURT 13784844 NIH MGC 184 Homo sapiens cDNA clone  
 IMAGE:30352015 5', mRNA sequence.  
 ACCESSION CB955620  
 VERSION CB955620.1 GI:30211738  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 331)  
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
 cDNA Library Preparation: CLOUTEC Laboratories, Inc.  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDCM150 row: h column: 08  
 High quality sequence stop: 319.  
 Location/Qualifiers  
 1..331  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30352015"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIH-MGC 184"  
 /note="Organ: Pooled-glandular; Vector: pDNR-LIB; Site: 1:  
 S11 (ggccatcggcc); Site 2: S11 (ggccgctggcc);  
 Library is oligo-dT primed and directionally cloned. cDNA  
 was prepared from a glandular pool of tissues from thyroid,  
 parathyroid, adrenal, cortex and pineal gland. 5' and 3'  
 adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor sequence:  
 5'-ATTGAGAGCGGAGCGCGCGACATG-dT(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.38  
 kb (range 0.60-3.5 kb). 15/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH-MGC Library."

BASE COUNT 93 a 79 c 55 g 103 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.017 Length: 331  
 Score: 110.00 Matches: 27  
 Percent Similarity: 61.82% Conservative: 7  
 Best Local Similarity: 49.09% Mismatches: 19  
 Query Match: 34.81% Indels: 2  
 DB: 14 Gaps: 2

US-10-079-754a-10 (1-58) x CB955620 (1-331)

Qy 1 MetLysIlePheHisPheValPheIleMetAlaLeuAlaMetIleArgAlaasp 20  
 Db 79 ATGAAGTTCCTGTCTTCCCTTCATCTTGCTTCATGATGTTCCATGATTTGAGACTAT 138

Qy 21 SerSerGluGluLysArgHisAsrGlySatGlySLysSHisHisArgGlyTyPrheGlnGln 40  
 Db 139 TCATCTGAAGAAATTTTGGGTAGAAATTTGAGAGATTGGATTATGGGTAT---GGCCCT 195

Qy 41 TygInProTyGInArgTyProLeuAsnTyProProAlaTyTyr 55  
 Db 196 TATCGAGCCAGTTCAGAACACACACTA---TACCACACACATAC 237

RESULT 8  
 LOCUS AA376695 341 bp mRNA linear EST 21-APR-1997  
 DEFINITION EST89151 Salivary gland Homo sapiens cDNA 5' end similar to  
 estherin, mRNA sequence.  
 ACCESSION AA376695  
 VERSION AA376695.1 GI:2029013  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 341)  
 AUTHORS Adams,M.D., Kirlavase,A.R., Fieschmann,R.D., Fulmer,R.A., Bull  
 C.J., Lee,N.H., Kirlavase,E.F., Weinstock,K.G., Gocayne,J.D., White  
 O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,  
 Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,  
 L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodex,A.,  
 Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,  
 Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merick,J.M.,  
 Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,  
 Small,K.V., Spriggs,T.A., Uterbeck,T.R., Weidman,J.F., Li,X.,  
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
 Dinke,D., Feng,D.-F., Fertle,A., Fischer,C., Hastings,G.A., He,M.W.,  
 Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,  
 Kunsch,C., Hungjun,J., Li,H., Weisner,P.S., Olsen,H., Raymond,L.,  
 Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon  
 M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Frazer,C.M. and  
 Venter,J.C.  
 Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)  
 MEDLINE 96026280  
 PUBMED 7566098

COMMENT Other\_ESTs: EST89150 THCT7891  
 Contact: Kirlavase, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: [arketlav@igr.org](mailto:arketlav@igr.org)  
 For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.

FEATURES  
 source  
 1..341  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="ATCC (inhost):181115"  
 /db\_xref="taxon:9606"  
 /dev\_stage="adult"  
 /clone\_lib="Salivary gland"  
 /note="Organ: salivary gland; Vector: pBluescript SK-;  
 Site 1: EcoRI; Site 2: XhoI"

BASE COUNT 99 a 82 c 53 g 106 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.0174 Length: 341  
 Score: 110.00 Matches: 27  
 Percent Similarity: 61.82% Conservative: 7  
 Best Local Similarity: 49.09% Mismatches: 19  
 Query Match: 34.81% Indels: 2  
 DB: 9 Gaps: 2

US-10-079-754A-10 (1-58) X AA376695 (1-341)

*Oy* MetLysIlePheLeIePreValIPheLIeMeAlAIdenILeuLametILEtArgLAap 20  
||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
*Dd* ATGAAGTCTCGTTGCTTGCCCTTCATCTGGCCTCATAGGTTCACAGTAGAGACTGAT 10  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
*Oy* SerSerGIunGLubSArghISARGLVSARGLYLSAHISHIRGSILTYrPhcIngIn 40  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
*Dd* TCATCTGAAGAAGAAATTGTCGTACAATTGGAAAGATTGGGTATGGGTAT--GCCT 16  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
*Oy* TYrGlNProTyrglNArgTYrProLeuAsnTYrProProAlatyr 55  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
*Dd* TATCAGCCAGTTCCAGAACACCACTH---TAACCACCAACCATAC 205  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 9	CB955686	LOCUS	DEFINITION
	CB955686	357 bp	mRNA
	AGENCOURT_13785438	NIH_MGC_184 Homo sapiens	linear EST 29-APR-2003
	IMAGE:30351237	5', mRNA sequence.	
	CCGCGCATG		

## FEATURES

**Source**

BASE COUNT	102 a	85 c	59 g	110̄ t	1 others
ORIGIN					

### Alignment Scores:

Pred. No.:

Score: \_\_\_\_\_

Percent Similarity  
Best Local Similarity

Query Match:

DB:

Length: 357

Matches: 27

Conservative: 7  
Mismatch: 19

```

Miscellaries: 13
Index: 2

```

AMCUB:	2
GADB:	2

1

US-10-079-754A-10 (1-58) X CB955686 (1-357)

```

Qy      1  MetLVysIePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAsp 20
Dy      76  ATAAAGTCTCTGTCTTTCCTTCATCTTGCTCTCAATGGTTCCATGATTTGAGACTGAT 13
Qy      21  SerSerGIuGIuIlyValrghIlsArGlylsArGlylsVlyshIshIlsArGIlyrPheGln 40
Dy      136  TCATCTGAAGAGAAAATTTTGGCTGACAAATGGAAAGATTCCGTTAAGGGAT---GGCCCT 19
Qy      41  TyrGlnProTyrGlnArGlyrProLeuShnTyrProProAlaTyr 55
Dy      193  TATCAAGCCAGTTCCAGAAACCACTA---TACCCACAACCACTAC 234

```

RESULT 10	AA376680	LOCUS	DEFINITION	LOCATION
AA376680	426 bp	mRNA	linear	EST 21-APR-1997
EST89133	Salivary gland Homo sapiens cDNA 5' end similar to			
scathearin, mRNA sequence.				
aaatccca				
aaatccca				

TITL

JOURN

MEDICAL

COMMENT



## Alignment Scores:

Pred. No.: 0.0206 Length: 426  
 Score: 110.00 Matches: 27  
 Percent Similarity: 61.82% Conservative: 7  
 Best Local Similarity: 49.09% Mismatches: 19  
 Query Match: 34.81% Indels: 2  
 DB: 9 Gaps: 2

US-10-079-754A-10 (1-58) x AA376680 (1-426)

QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAp 20  
 Db 61 ATGAAGTTCCTGCTCTTTCCTTCATCTTCGCTTCATGTTCCATGATTGAGCTGAT 120  
 QY 21 SerSerGluGluGlyAlaArgIleAlaGlyAlaGlyAlaGlyAlaGlyAlaGlyAla 40  
 Db 121 TCATCTGAAGAGAAATTTTGGCTGAAATTGAGAAATTCGTTATGGGTAT--GGCCCT 177  
 QY 41 TTYGlnProTYrGlnArgTYrProLeuAenTYrProProAlaTYr 55  
 Db 178 TATCAGCCAGTTCACAGAACACCACTA---TACCACACACCACTAC 219

## RESULT 11

CB957572

LOCUS

DEFINITION AGENCOURT 13785853 NIH MGC 184 Homo sapiens cDNA clone

IMAGE:30352366 5', mRNA sequence.

VERSION

CB957572.1 GI:30213689

KEYWORDS

EST.

SOURCE

ORGANISM

Human sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: c9abds@rmail.nih.gov

Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palokovits

cDNA Library Preparation: CLOUTEC Laboratories, Inc.

DNA sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: NDCM151 row: F column: 23

High quality sequence stop: 424.

Location/Qualifiers

1.441

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:30352366"

/lab\_host="DH10B (71 phage-resistant)"

/note="Organ: Pooled-glandular; Vector: pDNR-LIB; Site: 1;

Site1 (ggccatcggcc); Site 2: Site1 (ggccatcggcc);

Library is oligo-dT primed and directionally cloned. cDNA

was prepared from a glandular pool of tissues from thyroid,

parathyroid, adrenal, cortex and pineal gland. 5' and 3'

adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CACGCGCATTAATGCGC-3' and 3' adaptor sequence:

5'-ATTCTAGAGCGCGAGCGCGCGACACATG-dT(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.38

kb (range 0.60-3.5 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH MGC Library."

BASE COUNT 141 a 103 c 66 g 130 t 1 others

ORIGIN

## Alignment Scores:

Pred. No.: 0.0212 Length: 441  
 Score: 110.00 Matches: 27  
 Percent Similarity: 61.82% Conservative: 7  
 Best Local Similarity: 49.09% Mismatches: 19  
 Query Match: 34.81% Indels: 2  
 DB: 14 Gaps: 2

US-10-079-754A-10 (1-58) x CB957572 (1-441)

QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAp 20  
 Db 79 ATGAAGTTCCTGCTCTTTCCTTCATCTTCGCTTCATGTTCCATGATTGAGCTGAT 138  
 QY 21 SerSerGluGluGlyAlaArgIleAlaGlyAlaGlyAlaGlyAlaGlyAlaGlyAla 40  
 Db 139 TCATCTGAAGAGAAATTTTGGCTGAAATTGAGAAATTCGTTATGGGTAT--GGCCCT 195  
 QY 41 TTYGlnProTYrGlnArgTYrProLeuAenTYrProProAlaTYr 55  
 Db 136 TATCAGCCAGTTCACAGAACACCACTA---TACCACACACCACTAC 237

## RESULT 12

AA376682

LOCUS

DEFINITION AA376682 Salivary gland Homo sapiens cDNA 5' end similar to

stracherrin, mRNA sequence.

VERSION

AA376682.1 GI:2029000

KEYWORDS

EST.

SOURCE

ORGANISM

Human sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

7566098

Other ESTs: EST89134 EST89136 THC77891

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse

Location/Qualifiers

1.457

/organism="Homo sapiens"

BASE COUNT

ORIGIN



```
FT source 1..506
FT /db_xref="taxon:9606"
FT /mol_type="mRNA"
FT /organism="Homo sapiens"
FT /clone="DKFZp686F04245"
FT /clone_lib="686 (synonym: hicc3). Vector pSport1_Sfi; host
FT DH10B; sites SfiI + SfiIb"
FT /dev_stage="adult"
FT /tissue_type="CDNA-collection"
XX
SQ Sequence 506 BP; 152 A; 114 C; 72 G; 168 T; 0 other;

Alignment Scores:
Pred. No.: 0.0235 Length: 506
Score: 110.00 Matches: 27
Percent Similarity: 61.82% Conservative: 7
Best Local Similarity: 49.09% Mismatches: 19
Query Match: 34.81% Indels: 2
DB: 2 Gaps: 2

US-10-079-754a-10 (1-58) x HSM075180 (1-506)

QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAsp 20
Db 77 ATGAAGTTCCTGCTTCTTGCCTTCATCTTGCTCATGCTTCATGCTTCATGCTGAT 136
QY 21 SerSerGluGluLysArgHisArgLysArgLysHisArgGlyTyrPheGlnGln 40
Db 137 TCATCTGAAGAGAAATTTTGGCTGAGATTGGAAGATTGGCTTATGGGAT---GGCCCT 193
QY 41 TyrGlnProTyrGlnArgTyrProLeuAsnTyrProProAlaTyr 55
Db 194 TATCAGCCAGTTCAGAACACACACTA---TACCACACACCATAC 235

RESULT 15
LOCUS BG188342 517 bp mRNA linear EST 21-APR-2001
DEFINITION RST17360 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG188342
VERSION BG188342.1 GI:13710029
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 517)
Harrington,J.U., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Boozar,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
11329013
JOURNAL MEDLINE
PUBMED 11329013
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 517.
Location/Qualifiers
1..517
FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries Using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
```

cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 171 a 110 c 73 g 163 t

## ALIGNMENT SCORES:

Pred. No.: 0.0239 Length: 517  
Score: 110.00 Matches: 27  
Percent Similarity: 61.82% Conservative: 7  
Best Local Similarity: 49.09% Mismatches: 19  
Query Match: 34.81% Indels: 2  
DB: 10 Gaps: 2

US-10-079-754a-10 (1-58) x BG188342 (1-517)

```
QY 1 MetLysIlePheIlePheValPheIleMetAlaLeuIleLeuAlaMetIleArgAlaAsp 20
Db 19 ATGAAGTTCCTGCTTCTTGCCTTCATCTTGCTCATGCTTCATGCTTCATGCTGAT 78
QY 21 SerSerGluGluLysArgHisArgLysArgLysHisArgGlyTyrPheGlnGln 40
Db 79 TCATCTGAAGAGAAATTTTGGCTGAGATTGGAAGATTGGCTTATGGGAT---GGCCCT 135
QY 41 TyrGlnProTyrGlnArgTyrProLeuAsnTyrProProAlaTyr 55
Db 136 TATCAGCCAGTTCAGAACACACACTA---TACCACACACCATAC 177
```

Search completed: December 13, 2003, 04:23:25  
Job time : 1495 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2003, 01:04:32 ; Search time 2599 Seconds

(without alignments)  
9507.278 Million cell updates/sec

Title: US-10-079-754A-4

Perfect score: 604  
Sequence: 1 gaagatcttcagttcctata.....tatcaagcataaaaaaaa 604

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_ov.\*  
5: gb\_ov.\*  
6: gb\_ov.\*  
7: gb\_ph.\*  
8: gb\_ph.\*  
9: gb\_ph.\*  
10: gb\_ph.\*  
11: gb\_ph.\*  
12: gb\_ph.\*  
13: gb\_ph.\*  
14: gb\_ph.\*  
15: gb\_ph.\*  
16: gb\_ph.\*  
17: gb\_ph.\*  
18: gb\_ph.\*  
19: gb\_ph.\*  
20: gb\_ph.\*  
21: gb\_ph.\*  
22: gb\_ph.\*  
23: gb\_ph.\*  
24: gb\_ph.\*  
25: gb\_ph.\*  
26: gb\_ph.\*  
27: gb\_ph.\*  
28: gb\_ph.\*  
29: gb\_ph.\*  
30: gb\_ph.\*  
31: gb\_ph.\*  
32: gb\_ph.\*  
33: gb\_ph.\*  
34: gb\_ph.\*  
35: gb\_ph.\*  
36: gb\_ph.\*  
37: gb\_ph.\*  
38: gb\_ph.\*  
39: gb\_ph.\*  
40: gb\_ph.\*  
41: gb\_ph.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141	23.3	195534	2 AC134173	AC134173 Bos tauru
2	77	12.7	566	2 AC134173	AC134173 Bos tauru
3	75.4	12.5	357	6 BD071675	BD071675 Secreted
4	70.4	11.7	438	6 AX588147	AX588147 Sequence
5	70.4	11.7	438	6 AX588483	AX588483 Sequence
6	70.4	11.7	438	6 AX616321	AX616321 Sequence
7	68.2	11.3	524	9 HMMHS2X	M26665 Human histra
8	68.2	11.3	558	9 BC009791	BC009791 Homo sapi
9	67.6	11.2	542	9 HMMSTRNA	M18371 Human stach
10	67.6	11.2	552	9 HMMSTRNA	M18371 Human stach
11	67.6	11.2	1584	9 AK092678	AK092678 Homo sapi
12	67.2	11.1	203	6 BD071749	BD071749 Secreted
13	66.6	11.0	486	4 HMMHS1X	M26664 Human histra
14	65.6	10.9	486	4 AY154893	AY154893 Bos tauru
15	62.6	10.4	491	9 HMMHRPA	M18372 Human histra
16	60.8	10.1	516	6 BD071674	BD071674 Secreted
17	60.4	10.0	110000	2 AC106950_1	Continuation (2 of
18	60.4	10.0	231767	2 AC094531	Continuation (2 of
19	60.4	10.0	240446	2 AC131219	AC131219 Rattus no
20	59.2	9.8	7550	9 HMMHS102	L04132 Human histra
21	59.2	9.8	141568	9 AC063956	AC063956 Homo sapi
22	59.2	9.8	185969	2 AC069037	AC069037 Homo sapi
23	56.6	9.4	143842	2 AC134934	AC134934 Bos tauru
24	56.6	9.4	195534	2 AC134173	AC134173 Bos tauru
25	55.8	9.2	176	6 BD071751	BD071751 Secreted
26	55.6	9.2	218675	2 AC015847	AC015847 Homo sapi
27	55.6	9.1	249943	3 AC014823	AC014823 Plasmodiu
28	54.6	9.0	49306	3 AC115606	AC115606 Dictyoste
29	54.4	9.0	253001	3 AC014834	AC014834 Plasmodiu
30	54.2	9.0	154071	3 AC115598	AC115598 Dictyoste
31	54.2	9.0	250823	3 AC014821	AC014821 Plasmodiu
32	53.6	8.9	13449	6 AX346287	AX346287 Sequence
33	53.6	8.9	250022	3 AC014824	AC014824 Plasmodiu
34	52.4	8.7	82139	3 AC115684	AC115684 Dictyoste
35	52	8.6	186834	3 BX004887	BX004887 Zebrafish
36	52	8.6	349980	6 AX344567	AX344567 Sequence
37	51.8	8.6	254436	3 AC014827	AC014827 Plasmodiu
38	51.6	8.5	890	3 EHAPC	X76903 E.histolyti
39	51.6	8.5	7736	3 AF310889	AF310889 Dictyoste
40	51.6	8.5	36148	3 AC116100	AC116100 Dictyoste
41	51.6	8.5	110000	2 PFMALBP1_10	Continuation (11 o
42	51.4	8.5	161549	2 AC024676	AC024676 Homo sapi
43	51.2	8.5	6255	6 AX345863	AX345863 Sequence
44	51	8.4	12507	6 AX345200	AX345200 Sequence
45	51	8.4	12578	6 AX251413	AX251413 Sequence

# ALIGNMENTS

RESULT 1  
AC134173/c  
LOCUS  
DEFINITION  
AC134173  
Bos taurus clone RP42-254113, WORKING DRAFT SEQUENCE, 26 unordered  
pieces  
AC134173  
AC134173.1 GI:23306007  
VERSION  
KEYWORDS  
HTG, HTGS PHASE1, HTGS\_DRAFT.  
SOURCE  
Bos taurus (cow)  
ORGANISM  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE  
1 (bases 1 to 195534)  
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Pred. No. is the number of results predicted by chance to have a

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.bhsc.bcm.tmc.edu/docs/genbank\\_drift\\_data.html](http://www.bhsc.bcm.tmc.edu/docs/genbank_drift_data.html))  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 configs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the configs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES	Source
1	2358: contig of 2358 bp in length
2359	gap of unknown length
2459	contig of 3161 bp in length
5620	gap of unknown length
5720	contig of 2591 bp in length
8311	gap of unknown length
8411	contig of 2592 bp in length
11003	gap of unknown length
11103	contig of 3284 bp in length
14387	gap of unknown length
14487	contig of 2663 bp in length
17150	gap of unknown length
17250	contig of 4165 bp in length
21415	gap of unknown length
25151	contig of 3657 bp in length
25172	gap of unknown length
25272	contig of 5374 bp in length
30646	gap of unknown length
30746	contig of 5437 bp in length
36182	gap of unknown length
36183	contig of 5674 bp in length
36283	gap of unknown length
41957	contig of 6038 bp in length
42057	gap of unknown length
48095	gap of unknown length
48195	contig of 5020 bp in length
53215	gap of unknown length
53315	contig of 5535 bp in length
58850	gap of unknown length
58949	contig of 7195 bp in length
66145	gap of unknown length
66245	contig of 6789 bp in length
73033	gap of unknown length
73134	contig of 6897 bp in length
80031	gap of unknown length
80131	contig of 7631 bp in length
87762	gap of unknown length
87862	contig of 7865 bp in length
95727	gap of unknown length
95827	contig of 9228 bp in length
105055	gap of unknown length
105155	contig of 11562 bp in length
116717	gap of unknown length
116817	contig of 11477 bp in length
118294	gap of unknown length
128394	contig of 9747 bp in length
138141	gap of unknown length
138241	contig of 12447 bp in length
150688	gap of unknown length
150788	contig of 15627 bp in length
166414	gap of unknown length
166514	contig of 29020 bp in length
166515	gap of unknown length
195534	contig of 29020 bp in length
Location/Qualifiers	
1..195534	
/organism="Bos taurus"	
/mol_type="genomic DNA"	
/db_xref="taxon:9913"	
/clone="RP42-254113"	
60952 a 35784 c 36143 g 60048 t 2607 others	

-----  
\* NOTE: Estimated insert size may differ from sequence length

Query Match	23.3%	Score 141;	DB 2;	Length 195534;
Best Local Similarity	100.0%	Pred. No. 2.3e-17;		
Matches 141;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy 1 GAAGATTTTCACTCTATATAGATCTCATPACTGATGTAATTACAAACAAATGAA 60  
Db 47365 GAAGATTTTCACTCTATATAGATCTCATPACTGATGTAATTACAAACAAATGAA 47306  
Qy 61 GGATTTCAAGTATTTAAACACAGCAGTTTCTAGCAAGAACATCTCTGAGCATCG 120  
Db 47305 GGATTTCAAGTATTTAAACACAGCAGTTTCTAGCAAGAACATCTCTGAGCATCG 47246  
Qy 121 AATTTCATCTTCATGACTG 141  
Db 47245 AATTTCATCTTCATGACTG 47225

RESULT 2  
LOCUS BC017835 566 bp mRNA linear PRI 06-DEC-2001  
DEFINITION Homo sapiens, histatin 1, clone MGC:22502 IMAGE:4289874, mRNA,  
complete cds.  
ACCESSION BC017835  
VERSION BC017835.1 GI:17389614  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE Strausberg, R.  
JOURNAL Submitted (03-DEC-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: CLONTECH  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcdepax11.stanford.edu](mailto:mcdepax11.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

FEATURES  
source  
1..566  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:22502 IMAGE:4289874"  
/issue\_type="Skeletal Muscle"  
/clone\_lib="NIH\_MGC\_81"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
68..241  
/codon\_start=1  
/product="histatin 1"  
/protein\_id="AAH17835.1"  
/db\_xref="GI:17389615"  
/translation="MKFVFALVALIMISADSHKHHGRRKFKHSHREFP  
FYGDYSNVLNLYN"  
BASE COUNT 192 a 102 c 84 g 188 t  
ORIGIN

Query Match 12.7%; Score 77; DB 9; Length 566;  
Best Local Similarity 59.8%; Pred. No. 7.1e-05;

Matches 150; Conservative 0; Mismatches 95; Indels 6; Gaps 1;  
Qy 120 GAATTCATCTTTCATGACGTGACCCCAATATGAAGATCTTATCTTGTTCAT 179  
Db 34 GACTCTCTCTTGTAGTAAGAGACTCAGCACTATGAGTTTGTCTTGTTCAT 93  
Qy 180 TATGCTCTCATCTTACGATGATGATGATGATGATGATGATGATGATGATGAT 239  
Db 94 CTGGCTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 153  
Qy 240 ACGGAAAAACATCATAGAGATATTTTCAACAATACAGCATATCAAGATTCAC 299  
Db 154 TAGAAGAAATTCAT-----GAAAGCATCATTCACATGAGATTTTCATTTTANG 207  
Qy 300 AATATCTCTCTGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 359  
Db 208 GGACTATGATGATCAATATATATATGACATGATGATGATGATGATGATGATGAT 267  
Qy 360 TAGAGAGATTT 370  
Db 268 TATAGAGTTT 278

RESULT 3  
LOCUS BD071675 357 bp DNA linear PAT 27-AUG-2002  
DEFINITION Secreted expressed sequence tags (ESTs).  
ACCESSION BD071675  
VERSION BD071675.1 GI:22617278  
KEYWORDS JP 2001519667-A/485.  
SOURCE unidentified  
ORGANISM unidentified  
unclassified.

REFERENCE  
AUTHORS 1 (bases 1 to 357)  
TITLE Jacoby, K., McCoy, J. M., Lavallie, E. R., Racie, L. A., Merberg, D.,  
Trecy, M., Spaulding, V., and Agostino, M. J.  
JOURNAL Secreted expressed sequence tags (ESTs)  
Patent: JP 2001519667-A 485 23-OCT-2001;  
GENETICS INSTITUTE INC  
COMMENT OS Unidentified  
PN JP 2001519667-A/485  
PD 23-OCT-2001  
PF 10-APR-1998 JP 1998543069  
PR 10-APR-1997 US 08/838821  
PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI  
DAVID MERBERG,  
PI MAURICE TRECAY, VIRKI SPAULDING, MICHAEL J AGOSTINO PC  
C12N15/12,C12N5/10,C07K14/47,C12Q1/68,A61K38/17 CC Strandedness:  
Double;  
CC Topology: linear;  
CC Secreted expressed sequence tags (ESTs)  
FH Key Location/Qualifiers  
FT source 1..357  
FT /organism="Unidentified".

FEATURES  
source  
1..357  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 101 a 74 c 65 g 117 t  
ORIGIN  
Query Match 12.5%; Score 75.4; DB 6; Length 357;  
Best Local Similarity 59.4%; Pred. No. 0.00015;  
Matches 149; Conservative 0; Mismatches 96; Indels 6; Gaps 1;  
Qy 120 GAATTCATCTTTCATGACGTGACCCCAATATGAAGATCTTATCTTGTTCAT 179  
Db 48 GACTCTCTCTTGTAGTAAGAGACTCAGCACTATGAGTTTGTCTTGTTCAT 107  
Qy 180 TATGCTCTCATCTTACGATGATGATGATGATGATGATGATGATGATGATGAT 239  
Db 108 CTGGCTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 167

QY 240 ACGGAAAAACATAGAGATATTTTCAACATACAGCATATCAAGATTCACCT 299  
Db 168 TAGAGAAAAATTCAT-----GAAAAGCATCATTCACATCGAATTTTCATTATGG 221  
QY 300 AATATTCCTCCGCGATTCATTTCTTAAATGCTGCTTAGTACTACAGACATGAT 359  
Db 222 GGAAGTGGATCAATATTCATATGACAAATTCATATTCCTTAGTAAATCATGCGGATGAT 281  
QY 360 TAGAGATTT 370  
Db 282 TATAGAGTTT 292

RESULT 4  
AX588147  
LOCUS AX588147 438 bp DNA linear PAT 24-JAN-2003  
DEFINITION Sequence 22 from Patent WO02083898.  
ACCESSION AX588147  
VERSION AX588147.1 GI:27899822  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
Bejanin, S., Tanaka, H., Dumas Milne Edwards, J.B., Jobert, S. and  
Giordano, J.Y.  
TITLE Full-length human cdnas encoding potentially secreted proteins  
JOURNAL Patent: WO 02083898-A 22 24-OCT-2002;  
GENSET (FR)

FEATURES  
source location/Qualifiers  
1..438  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
84..320  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD61372.1"  
/db\_xref="GI:27899823"  
/translation="MKFVPALVLAISMISADSHKRRHGRRKFEKHSHYHTL  
LPLFEBSKSNANEKHNLTYLTCFRILAFSIVT"  
84..140  
/note="Von Heijne matrix score 7.64030745849671 seq  
ALVLAISMISADSHKRRHGRRKFEKHSHYHTL  
polysig\_peptide  
polysig\_signal 397..402  
polysig\_site 421..438  
BASE COUNT 153 a 80 c 60 g 143 t 2 others  
ORIGIN

Query Match 11.7%; Score 70.4; DB 6; Length 438;  
Best Local Similarity 69.9%; Pred. No. 0.0013;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACGCTCCACCAATATGAAATCTTTATCTTCTTCAT 179  
Db 50 GACTCTCTCTTGAGTAAAGACTCAGCCACATGAAGTTTGTCTTTGCTTTAGT 109  
QY 180 TATGGCTCTCATCTTACCATGATGATGATGATGATGATGATGATGATGATGATGAT 239  
Db 110 CTGGGCTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 169  
QY 240 ACGGAAAAACATCAT 255  
Db 170 TAGAGAAAAATTCAT 185

RESULT 5  
AX588483  
LOCUS AX588483 438 bp DNA linear PAT 24-JAN-2003  
DEFINITION Sequence 358 from Patent WO02083898.  
ACCESSION AX588483

VERSION AX588483.1 GI:27900156  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
Bejanin, S., Tanaka, H., Dumas Milne Edwards, J.B., Jobert, S. and  
Giordano, J.Y.  
TITLE Full-length human cdnas encoding potentially secreted proteins  
JOURNAL Patent: WO 02083898-A 358 24-OCT-2002;  
GENSET (FR)

FEATURES  
source location/Qualifiers  
1..438  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
84..320  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD61539.1"  
/db\_xref="GI:27900157"  
/translation="MKFVPALVLAISMISADSHKRRHGRRKFEKHSHYHTL  
LPLFEBSKSNANEKHNLTYLTCFRILAFSIVT"  
84..140  
/note="Von Heijne matrix score 7.59999990463257 seq  
ALVLAISMISADSHKRRHGRRKFEKHSHYHTL  
polysig\_peptide  
polysig\_signal 397..402  
polysig\_site 421..438  
BASE COUNT 151 a 80 c 60 g 145 t 2 others  
ORIGIN

Query Match 11.7%; Score 70.4; DB 6; Length 438;  
Best Local Similarity 69.9%; Pred. No. 0.0013;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACGCTCCACCAATATGAAATCTTTATCTTCTTCAT 179  
Db 50 GACTCTCTCTTGAGTAAAGACTCAGCCACATGAAGTTTGTCTTTGCTTTAGT 109  
QY 180 TATGGCTCTCATCTTACCATGATGATGATGATGATGATGATGATGATGATGATGAT 239  
Db 110 CTGGGCTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 169  
QY 240 ACGGAAAAACATCAT 255  
Db 170 TAGAGAAAAATTCAT 185

RESULT 6  
AX616321  
LOCUS AX616321 438 bp DNA linear PAT 20-FEB-2003  
DEFINITION Sequence 5 from Patent WO02094864.  
ACCESSION AX616321  
VERSION AX616321.1 GI:28447365  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
Bejanin, S. and Tanaka, H.  
TITLE Human cdnas and proteins and uses thereof  
JOURNAL Patent: WO 02094864-A 5 28-NOV-2002;  
GENSET (FR)

FEATURES  
source location/Qualifiers  
1..438  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
1..83  
84..320  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD67625.1"



```

/db_xref="GI:28447366"
/translation="MKFFVFAVLALMTSMISADSHKRRHGHGRKFKHSHSYHTL
LPLEESSKSNANKEHYNLVTLCPRIAFISYI"
3'UTR
polya_signal 318 .438
polya_site 397 .402
polya_site 423 .438
BASE COUNT 155 a 80 c 60 g 143 t
ORIGIN

Query Match 11.7%; Score 70.4; DB 6; Length 438;
Best Local Similarity 69.9%; Pred. No. 0.0013;
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTGACTGAGCTCCACCAATATGAGATCTTATCTTGTCTTCAT 179
DB 50 GACTCTCTCTGATGATAAAGAGCTCAGCAACTATGAGTTTGTCTTGTCTTACT 109
QY 180 TATGGCTCTCATCTGAGCCATGATTAAGAGCTGATTCATCTGAAGAAAGCTCAGAGAA 239
DB 110 CTGGCTCTCATGATTTCCATGATTAAGAGCTGATTCATGAAAGAGCATCATGGTA 169
QY 240 ACGGAAAAACATCAT 255
DB 170 TAGAAGAAAATTCAT 185

RESULT 7
HUMHIS2X 524 bp mRNA linear PRI 31-DEC-1994
DEFINITION Human histatin 2 (HIS2) mRNA, complete cds.
ACCESSION M26665
VERSION M26665.1 GI:292145
KEYWORDS histatin 2.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 524)
Sabatini, L.M. and Azen, E.A.
Histatins, a family of salivary histidine-rich proteins, are
encoded by at least two loci (HIS1 and HIS2)
Biochem. Biophys. Res. Commun. 160 (2), 495-502 (1989)
JOURNAL MEDLINE 89246491
PUBMED 2719677
COMMENT Original source text: Homo sapiens parotid gland cDNA to mRNA.
FEATURES
source location/Qualifiers
1..524
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="parotid gland"
1..524
/gene="HIS2"
2..227
/codon_start=1
/gene="HIS2"
/product="histatin 2"
/protein_id="AAAS8646.1"
/db_xref="GI:292146"
/translation="MKFFVFAVLALMTSMISADSHKRRHGHGRKFKHSHRGYR
SNLYIDN"
polya_signal 506 .511
/gene="HIS2"
BASE COUNT 157 a 100 c 87 g 180 t
ORIGIN

Query Match 11.3%; Score 68.2; DB 9; Length 524;
Best Local Similarity 65.4%; Pred. No. 0.0033;
Matches 100; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTGACTGAGCTCCACCAATATGAGATCTTATCTTGTCTTCAT 179
DB 38 GATTCCTCTGAGTAAAGAGCTCAGCAACTATGAGTTTGTCTTGTCTTAT 97

```

```

QY 180 TATGGCTCTCATCTGAGCCATGATTAAGAGCTGATTCATCTGAAGAAAGCTCAGAGAA 239
DB 98 CTGGCTCTCATGCTTTCCATGATGAGCTGAGCTGATTCATGCAAGAGCATCATGGTA 157
QY 240 ACGGAAAAACATCATGAGGATATTTTCAACA 272
DB 158 TAAAGAAAATTCATGAAAGAGCATCATTCACA 190

RESULT 8
LOCUS BC009791 558 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, histatin 3, clone MGC:13578 IMAGE:4293405, mRNA,
complete cds.
ACCESSION BC009791
VERSION BC009791.1 GI:14602560
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 558)
Strausberg, R.
Direct Submission
Submitted (02-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbio.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNLT at: http://image.llnl.gov
Series: IPAL Plate: 19 Row: 0 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4557652.
location/Qualifiers
1..558
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="locusID:3347"
/db_xref="taxon:9606"
/clone="MGC:13578 IMAGE:4293405"
/tissue_type="Skeletal Muscle"
/clone_id="NIH MGC_81"
/lab_host="DH10B"
/notes="Vector: pDNR-LIB"
76..231
/codon_start=1
/product="histatin 3"
/protein_id="AAH09791.1"
/db_xref="GI:14602561"
/translation="MKFFVFAVLALMTSMISADSHKRRHGHGRKFKHSHRGYR
SNLYIDN"
BASE COUNT 187 a 102 c 91 g 178 t
ORIGIN

Query Match 11.3%; Score 68.2; DB 9; Length 558;
Best Local Similarity 65.4%; Pred. No. 0.0033;
Matches 100; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTGACTGAGCTCCACCAATATGAGATCTTATCTTGTCTTCAT 179
DB 111 GATTCCTCTGAGTAAAGAGCTCAGCAACTATGAGTTTGTCTTGTCTTAT 97

```

Db 42 GATTCCTCTCTGAGTAAAGACTCAGCAACTGAGATTTTGTCTTAAAT 101  
Qy 180 TATGGCTCTATCTAGCATGATTAAGCTGATTCATCTGAAGAAAGTCACAGGAA 239  
Db 102 CTGGCTCTATGCTTTCATGACTGAGCTGATGATGCAAGAGACATCTGGGTA 161  
Qy 240 ACGGAAAAAATCATGATGAGATATTTTCAACA 272  
Db 162 TAAAGAAAATTCATGAAAAGCATCTACCA 194

RESULT 9  
HUMSTRNA 542 bp mRNA linear PRI 13-JAN-1995  
LOCUS Human statherin mRNA, complete cds.  
DEFINITION M18371.1 GI:338610  
ACCESSION M18371.1 GI:338610  
VERSION M18371.1 GI:338610  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE 1 (bases 1 to 542)  
AUTHORS Dickinson,D.P., Ridall,A.L. and Levine,M.V.  
Human submandibular gland statherin and basic histidine-rich  
peptide are encoded by highly abundant mRNA's derived from a common  
ancestral sequence  
JOURNAL Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)  
MEDLINE 88106506  
COMMENT Original source text: Human female submandibular gland, cDNA to  
mRNA, clone PBRHSMF98.2.  
Draft entry and computer readable copy of sequence [1] kindly  
provided by D.P. Dickinson 21-MAR-1988.  
Location/Qualifiers  
source 1..542  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="4q11-q13"  
1..542  
/gene="STATH"  
57..245  
/gene="STATH"  
/note="statherin precursor"  
/codon\_start=1  
/protein\_id="AAA60600.1"  
/db\_xref="GI:338611"  
/db\_xref="GDB:G00-120-391"  
/translation="MKFLVFAFLALVMWIMIGADSESEKFLRIRIGFGYGYQYQYV  
EQLYRQYQYQYQYTF"  
57..113  
sig\_peptide  
/gene="STATH"  
/note="statherin signal peptide"  
114..242  
/gene="STATH"  
/product="statherin"  
170 a 117 c 76 g 179 t  
BASE COUNT 262 bp upstream of PstI site; chromosome 4q11-q13.  
ORIGIN  
Query Match 11.2%; Score 67.6; DB 9; Length 542;  
Best Local Similarity 61.2%; Pred. No. 0.0042;  
Matches 147; Conservative 0; Mismatches 84; Indels 9; Gaps 2;

Qy 140 GGAATCCAAATATGAGATCTTTATCTTTGCTTCATTAAGCTCTCATCTACCA 199  
Db 43 GAACCCAGCCCAATATGAGATCTTTGCTTCATTAAGCTCTCATCTACCA 102  
Qy 200 TGATTAGAGCTGATTCATCTGAAGAAAGCTCAGAGAAACGAAAAACATATAGAG 259  
Db 103 TGATTGAGCTGATTCATCTGAAGAAATTTTGGTAGAATTCGAGATTCGGTATG 162  
Qy 260 GATA-----TTTCAACAATATACAGCATATCAAGATATCACTAAATATCTCTG 313

Db 163 GGTATGACCTTATCAGCAATTCAGAAACCAACTATACCAACATACCAACAC 222  
Qy 314 CGT---ATCCATTTCTTAAATGCTGCTTAGTAATACAGACATGATTAAGATTT 370  
Db 223 AATACCAACAATATACCTTTTATATATCATGACTGACGACGATGATTTAGAGCTT 282

RESULT 10  
HUMSTRNA 552 bp mRNA linear PRI 13-JAN-1995  
LOCUS Human statherin mRNA, complete cds.  
DEFINITION M18078.1 GI:338507  
ACCESSION M18078.1 GI:338507  
VERSION M18078.1 GI:338507  
KEYWORDS regulatory protein; statherin.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 552)  
AUTHORS Sabatini,L.M., Carllock,L.R., Johnson,G.W. and Azen,E.A.  
CDNA cloning and chromosomal localization (4q11-13) of a gene for  
statherin, a regulator of calcium in saliva  
JOURNAL Am. J. Hum. Genet. 41 (6), 1048-1060 (1987)  
MEDLINE 88074310  
COMMENT Original source text: Human parotid gland, cDNA to mRNA, clone  
H772B.  
Draft entry and computer-readable sequence [1] kindly submitted by  
L.Sabatini 19-JAN-1988.  
Location/Qualifiers  
source 1..552  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="4q11-q13"  
1..552  
/gene="STATH"  
73..261  
/gene="STATH"  
/note="statherin protein"  
/codon\_start=1  
/protein\_id="AAA60594.1"  
/db\_xref="GI:338508"  
/db\_xref="GDB:G00-120-391"  
/translation="MKFLVFAFLALVMWIMIGADSESEKFLRIRIGFGYGYQYQYV  
EQLYRQYQYQYQYTF"  
73..129  
sig\_peptide  
/gene="STATH"  
/note="statherin signal peptide"  
114..242  
BASE COUNT 172 a 122 c 78 g 180 t  
ORIGIN 278 bp upstream of PstI site; chromosome 4q11-q13.  
Query Match 11.2%; Score 67.6; DB 9; Length 552;  
Best Local Similarity 61.2%; Pred. No. 0.0042;  
Matches 147; Conservative 0; Mismatches 84; Indels 9; Gaps 2;

Qy 140 GGAATCCAAATATGAGATCTTTATCTTTGCTTCATTAAGCTCTCATCTACCA 199  
Db 59 GAACCCAGCCCAATATGAGATCTTTGCTTCATTAAGCTCTCATCTACCA 118  
Qy 200 TGATTAGAGCTGATTCATCTGAAGAAAGCTCAGAGAAACGAAAAACATATAGAG 259  
Db 119 TGATTGAGCTGATTCATCTGAAGAAATTTTGGTAGAATTCGAGATTCGGTATG 178  
Qy 260 GATA-----TTTCAACAATATACAGCATATCAAGATATCACTAAATATCTCTG 313  
Db 179 GGTATGACCTTATCAGCAATTCAGAAACCAACTATACCAACATACCAACAC 238  
Qy 314 CGT---ATCCATTTCTTAAATGCTGCTTAGTAATACAGACATGATTAAGAGATTT 370  
Db 239 AATACCAACAATATACCTTTTATATATCATGACTGACGACGATGATTTAGAGCTT 298

RESULT 11  
AK092678 1584 bp mRNA linear PRI 15-JUL-2002  
LOCUS AK092678  
DEFINITION Homo sapiens CDNA FLJ35359 file, clone SALGL1000107, highly similar to CH-ROG PROTEIN.  
ACCESSION AK092678  
VERSION AK092678.1 GI:21753327  
KEYWORDS oligo cloning; file (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1  
AUTHORS Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K., Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R., Otsubu,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Magatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isogai,T.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1584)  
AUTHORS Isogai,T. and Yamamoto,J.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genom@cdhri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB) (supported by Japan Key Technology Center etc.); 5- & 3-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
FEATURES  
source location/Qualifiers  
1..1584  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="SALGL1000107"  
/issue\_type="salivary gland"  
/clone\_id="SALGL1"  
/note="Cloning vector: pWE18FLJ3"  
89..997  
/note="unnamed protein product"  
/protein\_id="BAC03943.1"  
/db\_xref="GI:21753328"  
/translation="MGDSSEWLKLPVDCKEHLKMKARLSGYEALKIFQIKDEKSP EWSKFLGILKKFVDSNAVVDLKGLEALVVENAHVAGKTGEVSVSVSVFQPK AKAKELGIEICLMYIEIKGEVQSEELKGLDNKPKIVACIETLRALSSEFGSKII LKPIIKVLPKLFESREKAVDEAKLIVETIRYRDLRPLOININSVOLKEEEN VKLPFSARPTPLFESQOELEKLEQOQSAGSDAGGDDGDEVQIDAYELLEAVEI LSKLPKMTKLRQNGKREKRPMSL"  
BASE COUNT 526 a 298 c 331 g 429 t  
ORIGIN  
Query Match 11.2%; Score 67.6; DB 9; Length 1584;  
Best Local Similarity 61.2%; Pred. No. 0.0036;  
Matches 147; Conservative 0; Mismatches 84; Indels 9; Gaps 2;

QY 260 GATA-----TTTCAACAATACAGCCATATCAAGATATCCATAATTACTCTCTG 313  
DB 1203 GGTATGGCCCTTATCAGCCAGTTCAGAAACACCACTATCCACCAACCATACCAACAC 1262  
QY 314 CGT--ATCCATTTCCTTAAATGCTGTGTGTAATCTACAGACATGATTAGAGATT 370  
DB 1263 AATACCAACAATATACCTTTTAAATATCATCATGATACAGACATGATTATGAGGCTT 1322  
RESULT 12  
BD071749 203 bp DNA linear PAT 27-AUG-2002  
LOCUS BD071749  
DEFINITION Secreted expressed sequence tags (SESTs).  
ACCESSION BD071749  
VERSION BD071749.1 GI:22617352  
KEYWORDS JP 2001519667-A/559.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 203)  
AUTHORS Jacobs,K., McCoy,J.M., Lavallie,E.R., Racie,L.A., Werberg,D., Treacy,M., Spaulding,V. and Agostino,M.J.  
TITLE Secreted expressed sequence tags (SESTs)  
JOURNAL Patent: JP 2001519667-A 559 23-OCT-2001;  
COMMENT GENETICS INSTITUTE INC  
OS Unidentified  
PN JP 2001519667-A/559  
PD 23-OCT-2001  
PF 10-APR-1998 JP 1998543069  
PR 10-APR-1997 US 08/838821  
PI KENNETH JACOB,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE, PI DAVID WERBERG,  
PI MAURICE TREACY,VIKKI SPAULDING,MICHAEL J AGOSTINO PC  
C1N15/12.C1N15/10.C07K14/47.C1201/68.A61K38/17 CC Strandedness:  
Double;  
CC Topology: linear;  
CC Secreted expressed sequence tags (SESTs)  
FH key  
FT source 1..203  
/organism="Unidentified".  
FEATURES  
source location/Qualifiers  
1..203  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 59 a 43 c 39 g 62 t  
ORIGIN  
Query Match 11.1%; Score 67.2; DB 6; Length 203;  
Best Local Similarity 68.4%; Pred. No. 0.0059;  
Matches 93; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 120 GAATTCATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 179  
DB 47 GATTCCTCTTGAAGTAAGATGATGATGATGATGATGATGATGATGATGATGAT 106  
QY 180 TATGCTCTCATCTCAGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 239  
DB 107 CTGGCTCTCATGCTTCCTTCATGATGATGATGATGATGATGATGATGATGATGAT 166  
QY 240 ACGGAAAAAATCATCT 255  
DB 167 TAAAGAAAAATTCAT 182  
RESULT 13  
HMHHSIX 480 bp mRNA linear PRI 31-DEC-1994  
LOCUS HMHHSIX  
DEFINITION Human histatin 1 (HIS1) mRNA, complete cds.  
ACCESSION M26664  
VERSION M26664.1 GI:292143  
KEYWORDS histatin 1.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 480)  
AUTHORS Sabatini, L.M. and Azen, E.A.  
TITLE Histatins, a family of salivary histidine-rich proteins, are  
encoded by at least two loci (HIS1 and HIS2)  
JOURNAL Biochem. Biophys. Res. Commun. 160 (2), 495-502 (1989)  
MEDLINE 89246491  
PUBMED 2719677  
COMMENT Original source text: Homo sapiens parotid gland cDNA to mRNA.  
FEATURES  
source  
location/Qualifiers  
1..480  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="parotid gland"  
1..480  
/gene="HIS1"  
10..183  
/gene="HIS1"  
/codon\_start=1  
/product="histatin 1"  
/protein\_id="AA58645.1"  
/db\_xref="GI:292144"  
/translation="MKFFVPAVLALMISMISADSHKHHGRRKFEKHHSHREFP  
FYDYGSNLYDN"  
462..467  
/gene="HIS1"  
BASE COUNT 152 a 85 c 75 g 168 t  
ORIGIN  
polya\_signal  
Query Match 11.0%; Score 66.6; DB 9; Length 480;  
Best Local Similarity 59.6%; Pred. No. 0.0067;  
Matches 133; Conservative 0; Mismatches 84; Indels 6; Gaps 1;  
QY 148 CCAAAATGAGAGCTTTATCTTGTCTTCATTATGGCTCTCATCCAGCATGATTAGA 207  
DB 4 CCAACTATGAGATTTTGTCTTGTCTTGTAGTCTGCTCATGATTCATGATTAGC 63  
QY 208 GCTGATTCATCTGAAGAGAAAGCTCACAGAGAAAGGAAAAATCATCTAGAGATATTTT 267  
DB 64 GCTGATTCATCTGAAGAGAAAGCTCACAGAGAAAGGAAAAATCATCTAGAGATATTTT 267  
QY 268 CACCAATACAGAGCATATCAACGATATCCACTAAATATTCCTCGCTATCCATTTCT 327  
DB 118 CACTATTCACATCGAGATTTCCATTTTATGGGACATATGATTAATTTCTATATGAC 177  
QY 328 TAAATGCTGCTTGAATCACTACAGACATGATTAGAGATTT 370  
DB 178 AATTGATATCTCTAGTATCATGCGGCATGATTATAGAGTTT 220  
RESULT 14  
AY154893 486 bp mRNA linear MAM 17-DEC-2002  
LOCUS Bos taurus statherin (STATH) mRNA, complete cds.  
DEFINITION  
ACCESSION AY154893  
VERSION AY154893.1 GI:27227454  
KEYWORDS  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Rijnkeels, M., Elhiteki, L., Miller, W. and Rosen, J.M.  
TITLE Multi-species comparative sequence analysis of the casein gene  
cluster region: A gene domain encoding epithelial secretory  
proteins  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 486)  
AUTHORS Rijnkeels, M. and Rosen, J.M.

TITLE Direct Submission  
JOURNAL Submitted (24-SEP-2002) Molecular and Cellular Biology, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
FEATURES  
source  
location/Qualifiers  
1..486  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/chromosome="6"  
/tissue\_type="salivary gland"  
1..486  
/gene="STATH"  
1..180  
/gene="STATH"  
/codon\_start=1  
/product="statherin"  
/protein\_id="AA085581.1"  
/db\_xref="GI:27227455"  
/translation="MKIFFPAPLALMVMIRKADSEEEHRLRFPNPFYYPNOGGYI  
PSYPAYPPYPYVQ"  
BASE COUNT 155 a 92 c 79 g 160 t  
ORIGIN  
Query Match 10.9%; Score 65.6; DB 4; Length 486;  
Best Local Similarity 62.8%; Pred. No. 0.01;  
Matches 123; Conservative 0; Mismatches 64; Indels 9; Gaps 1;  
QY 154 ATGAGATCTTTATCTTGTCTTCATTATGGCTCTCATCCATGATTAAGCTGAT 213  
DB 1 ATGAGATCTTTTCTTGTCTTCTTATATGAGCTCTATGTTCCATGATTAAGCTGAT 60  
QY 214 TCATCTGAAGAGAAAGCTCACAGAGAAAGGAAAAATCATCTAGAGATATTTTCA 273  
DB 61 TCATCTGAAGAGAGAAC-----ACCGCTTGAATTTAACCTGATTTTACCCA 111  
QY 274 TACAGCATATCAAGATATCCATTAATATTCCTCGGATTCATTTCTTAAAT 333  
DB 112 AACACAGAGTGAATATATTCCTAGTTATCCAGCATATCATATCATATTC 171  
QY 334 GCTGCTTGAATAC 349  
DB 172 GTCCAGTAGTACGCTG 187  
RESULT 15  
HUMBRPA 491 bp mRNA linear PRI 31-OCT-1994  
LOCUS Human histidine-rich protein mRNA, complete cds.  
DEFINITION  
ACCESSION M18372  
VERSION M18372.1 GI:179465  
KEYWORDS histidine-rich protein.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 491)  
AUTHORS Dickinson, D.P., Ridall, A.L. and Levine, M.J.  
TITLE Human submandibular gland statherin and basic histidine-rich  
peptide are encoded by highly abundant mRNA's derived from a common  
ancestral sequence  
JOURNAL Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)  
MEDLINE 88106506  
PUBMED 3426601  
COMMENT Original source text: Human female submandibular gland, cDNA to  
mRNA, clone pBRH5SF1B8.1.  
Draft entry and computer readable copy of sequence [1] kindly  
provided by D.P. Dickinson 21-MAR-1988.  
FEATURES  
source  
location/Qualifiers  
1..491  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="3q21-qter"

```

gene      1..491
           /gene="HRG"
CDS        38..193
           /gene="HRG"
           /note="precursor"
           /codon_start=1
           /product="basic histidine-rich protein"
           /protein_id="AA51830.1"
           /db_xref="GI:179466"
           /db_xref="GDB:G00-120-055"
           /translation="MKFVFALILALMSMTGADSHAKRHGYKRFHKHSHRGYR
SNLYLDN"

sig_peptide 38..94
            /gene="HRG"
            /note="Basic histidine-rich protein signal peptide"

mat_peptide 95..190
            /gene="HRG"
            /product="basic histidine-rich protein"

BASE COUNT 151 a      90 c      87 g     163 t
ORIGIN      Unreported.

Query Match      10.4%: Score 62.6; DB 9; Length 491;
Best Local Similarity 66.9%; Pred. No. 0.038;
Matches 89; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY      140 GGACTCAACCAATATGAGATCTTTATCTTTCATTATGAGCTCTCATCTAGCCA 199
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      24 GGATTCACCACTATGAGATTTTGTGTTTGTCTTAATCTTGCTCTCATGCTTCCA 83

QY      200 TGATTAGAGCTATTCATCTGAAGAAAGTCACAGAAACGAAAAACATCATAGAG 259
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      84 TGACTGAGAGCTATTCATGCAAGAGACATCATGGGTATTAAGAAAAATTCATGAAA 143

QY      260 GATATTTTCACA 272
        ||| ||| |||
DB      144 AGCATCATTCACA 156

```

Search completed: December 13, 2003, 02:29:18  
Job time : 2602 secs

**THIS PAGE BLANK (USPTO)**



PR	22-AUG-2000; 2000US-0644190.	
PR	27-OCT-2000; 2000US-0639146.	
XX		
XX	(GENE-) GENESIS RES & DEV CORP LTD.	
PA		
XX		
PI	Glenn M, Grigor MR, Molenaar AJ, Davis SR;	
XX		
DR	WPI: 2003-275306/27.	
DR	P-PSDB: ABG75794.	
XX		
PT	New histatin polynucleotides and polypeptides expressed in bovine	
PT	tissues, useful for treating bacterial or fungal infections in mammals,	
PT	e.g. infections of the oral cavity, vagina, urethra, ear or skin, or	
PT	systemic infections -	
PS	Claim 1; Page 12; 17pp; English.	
XX		
CC	The invention discloses isolated histatin polynucleotides and	
CC	polypeptides, which comprise sequences expressed in bovine. Histatins are	
CC	histidine rich and specific to the salivary secretions. They are believed	
CC	to function as part of the non-immune defence system, particularly in the	
CC	oral cavity and have promise as therapeutic agents in humans with oral	
CC	candidosis. Also disclosed are compositions comprising the histatin	
CC	polypeptide or polynucleotide, and at least one component consisting of	
CC	physiological or pharmaceutical carriers or immunostimulants. The	
CC	histatin polynucleotide, polypeptide, or the cosmetic composition	
CC	comprising the polypeptide, is useful for treating (e.g. gene therapy) a	
CC	disorder in a mammal, particularly microbial or fungal infections. The	
CC	histatin polynucleotide is also useful in genome mapping, physical	
CC	mapping or in the positional cloning of genes. Specifically, the	
CC	polypeptide or polynucleotide is useful for treating fungal or bacterial	
CC	infections of the oral cavity (e.g. dental caries, plaque or tartar),	
CC	vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis	
CC	caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating	
CC	systemic infections such as systemic Candida infection. The polypeptide	
CC	or polynucleotide is also useful for veterinary applications e.g. for	
CC	treating mastitis. The sequence presented is a cDNA encoding a bovine	
CC	histatin like polypeptide.	
XX		
SQ	Sequence 604 BP; 222 A; 112 C; 79 G; 191 T; 0 other;	
	Query Match	100.0%; Score 604; DB 25; Length 604;
	Best Local Similarity	100.0%; Pred. No. 1,2e-120;
	Matches 604; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	GAAGTATTTGAGTCTATATTAAGATCTCAACTGATGATATTAACAAAACAAATGA 60
DB	1	GAAAGATTTTGAATCTAAATTAATGAATCTCATACACTGATGATTAACAAAACAAATGA 60
QY	61	GGATTTCAAGGTAATTTAAACACAGCAGTTTTCTAGCAAGAACATCTCTGAAGCATG 120
DB	61	GGATTTCAAGGTAATTTAAACACAGAGTTTTCTAGCAAGAACATCTCTGAAGCATG 120
QY	121	AATTTCAATCTTTTCATGACTGAGCTGCACCAATATGAAGATCTTATCTTTGCTTCATT 180
DB	121	AATTTCAATCTTTTCATGACTGAGCTGCACCAATATGAAGATCTTATCTTTGCTTCATT 180
QY	181	ATGGCTCATATCCATGACCATATTAAGACCTGATTCATGGAAGAAACGTACAGGAA 240
DB	181	ATGGCTCATATCCATGACCATATTAAGACCTGATTCATGGAAGAAACGTACAGGAA 240
QY	241	CGGAAAAAACAATCATAAGAGATATTTTCAACACATACCGACATATCAACGATATCCACTA 300
DB	241	CGGAAAAAACAATCATAAGAGATATTTTCAACACATACCGACATATCAACGATATCCACTA 300
QY	301	AATTATCTCTGCGGTATCCATTTCTTAAATGCTGCTTAGTACTACAGACATGATT 360
DB	301	AATTATCTCTGCGGTATCCATTTCTTAAATGCTGCTTAGTACTACAGACATGATT 360
QY	361	AGAGAGATTTTTCACATGATTTTTCATCTTTCTGTGTGTTGAAGAACATCTTTCA 420
DB	361	AGAGAGATTTTTCACATGATTTTTCATCTTTCTGTGTGTTGAAGAACATCTTTCA 420

Qy	421	AATGAAATAAAA	CAAGAAAAA	AAAAAATTCAGTCAAGTAGTTGCACAACACATCTTGGAAATC	480
Db	421	AATGAAATAAAA	CAAGAAAAA	AAAAAATTCAGTCAAGTAGTTGCACAACACATCTTGGAAATC	480
Qy	481	AAATATCAATA	TTTTAAAA	CAATATATATGATAGTCTCTGAACATATGTAAATGGTTTCTAC	540
Db	481	AAATATCAATA	TTTTAAAA	CAATATATATGATAGTCTCTGAACATATGTAAATGGTTTCTAC	540
Qy	541	TTTTCTTTCTCTGTCATCTTACCATGCAATGCTTAATTAATTTGATCTATCAAGCATTA	AAAAA	600	
Db	541	TTTTCTTTCTCTGTCATCTTACCATGCAATGCTTAATTAATTTGATCTATCAAGCATTA	AAAAA	600	
Qy	601	AAAA	604		
Db	601	AAAA	604		

CC	RESULT 2
XX	ABX11388
ID	ABX11388 standard; cDNA; 505 BP.
AC	
XX	ABX11388;
DT	01-MAY-2003 (first entry)
XX	
DE	Bovine histatin like polypeptide, #1, cDNA.
XX	
KW	Cow; gene; ss; histatin; bovine; salivary secretion; oral cavity;
KW	non-immune defence system; oral candidosis; gene therapy;
KW	microbial infection; fungal infection; dental caries; plaque; tartar;
KW	cystic fibrosis; systemic infection; Candida infection; mastitis;
KW	fungicide; antibacterial.
XX	
OS	Bos taurus.
PH	
FT	Key Location/Qualifiers
FT	CDS 44..220
FT	/*tag= a
FT	/product= "Histatin like polypeptide"
XX	
PN	US2002164625-A1.
XX	
PD	07-NOV-2002.
XX	
PF	19-FEB-2002; 2002US-0079754.
XX	
PR	23-AUG-1999; 99US-150330P.
PR	29-OCT-1999; 99US-162701P.
PR	22-AUG-2000; 2000US-0644190.
PR	27-OCT-2000; 2000US-0699146.
XX	
PA	(GENE-) GENESIS RES & DEV CORP LTD.
XX	
PI	Glenn M, Grigor MR, Molenaar AJ, Davis SR;
XX	
DR	WPI; 2003-275306/27.
XX	
PT	P-PsDB; ABG75791.
XX	
PT	New histatin polynucleotides and polypeptides expressed in bovine
PT	tissues, useful for treating bacterial or fungal infections in mammals,
PT	e.g. infections of the oral cavity, vagina, urethra, ear or skin, or
PT	systemic infections -
XX	
PS	Claim 1; Page 11; 17pp; English.
XX	
CC	The invention discloses isolated histatin polynucleotides and
CC	polypeptides, which comprise sequences expressed in bovine. Histatins are
CC	histidine rich and specific to the salivary secretions. They are believed
CC	to function as part of the non-immune defence system, particularly in the
CC	oral cavity and have promise as therapeutic agents in humans with oral
CC	candidosis. Also disclosed are compositions comprising the histatin
CC	polypeptide or polynucleotide, and at least one component consisting of
CC	physiological or pharmaceutical carriers or immunostimulants. The



histatin polynucleotide, polypeptide, or the cosmetic composition comprising the polypeptide, is useful for treating (e.g. gene therapy) a disorder in a mammal, particularly microbial or fungal infections. The histatin polynucleotide is also useful in genome mapping, physical mapping or in the positional cloning of genes. Specifically, the polypeptide or polynucleotide is useful for treating fungal or bacterial infections of the oral cavity (e.g. dental caries, plaque or tartar), vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis caused by *Pseudomonas aeruginosa*), mucosa or eye, as well as for treating systemic infections such as systemic *Candida* infection. The polypeptide or polynucleotide is also useful for veterinary applications e.g. for treating mastitis. The sequence presented is a cDNA encoding a bovine histatin like polypeptide.

Sequence 505 BP; 188 A; 96 C; 67 G; 154 T; 0 other;

Query Match 77.1%; Score 465.4; DB 25; Length 505;  
Best Local Similarity 99.8%; Pred. No. 7.2e-91;  
Matches 466; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 138 CTGACCTCCACCAATATGAGATCTTTCTTTGCTTCAATTATGGCTCTCATCCTAC 197
DB 28 CAGGACTCCACCAATATGAGATCTTTCTTTGCTTCAATTATGGCTCTCATCCTAC 87
QY 198 CATGATTAGAGCTGATTCATCTGAAGAGAAAGCTCAAGAGAAACGAAAAAATCATATG 257
DB 88 CATGATTAGAGCTGATTCATCTGAAGAGAAAGCTCAAGAGAAACGAAAAAATCATATG 147
QY 258 AGGATATTTTCAACATATACCATATCAAGATATCCATTAATTTATCTCTGCGTA 317
DB 148 AGGATATTTTCAACATATACCATATCAAGATATCCATTAATTTATCTCTGCGTA 207
QY 318 TCCATTTCTTAAATGCTGCTTAGTAATACAGGACATGATTAAGAGATTTTTCACA 377
DB 208 TCCATTTCTTAAATGCTGCTTAGTAATACAGGACATGATTAAGAGATTTTTCACA 267
QY 378 TCAATTTTCTTAAATGCTGCTTAGTAATACAGGACATGATTAAGAGATTTTTCACA 437
DB 268 TCAATTTTCTTAAATGCTGCTTAGTAATACAGGACATGATTAAGAGATTTTTCACA 327
QY 438 AAAAAAATCAGTCAAGTATGTCACACACATCTTGAATCAATATCAATATTTTAA 497
DB 328 AAAAAAATCAGTCAAGTATGTCACACACATCTTGAATCAATATCAATATTTTAA 387
QY 498 AACATATATATATGCTCTGACATGATTAATGCTTCTTCTTCTCTGCTAC 557
DB 388 AACATATATATATGCTCTGACATGATTAATGCTTCTTCTTCTCTGCTAC 447
QY 558 TTACCATGATGCTTAAATTAATGATCTATCAAGCATATAAAAAAAA 604
DB 448 TTACCATGATGCTTAAATTAATGATCTATCAAGCATATAAAAAAAA 494

```

RESULT 3  
ABX11389  
ID ABX11389 standard; cDNA; 585 BP.  
XX  
AC ABX11389;  
XX  
DT 01-MAY-2003 (first entry)  
XX  
DE Bovine histatin like polypeptide, #2, cDNA.  
XX  
KM Cow; gene; ss; histatin; bovine; salivary secretion; oral cavity;  
KM non-immune defence system; oral candidosis; gene therapy;  
KM microbial infection; fungal infection; dental caries; tartar;  
KM cystic fibrosis; systemic infection; *Candida* infection; mastitis;  
KM fungicide; antibacterial.  
XX  
OS Bos taurus.  
XX  
XX Key  
FH CDS Location/Qualifiers  
FT CDS 149..325

/\*tag= a  
/product= "Histatin like polypeptide"

US2002164625-A1.  
07-NOV-2002.  
19-FEB-2002; 2002US-0079754.  
23-AUG-1999; 99US-150330P.  
29-OCT-1999; 99US-162701P.  
22-AUG-2000; 2000US-0644190.  
27-OCT-2000; 2000US-0699146.

(GENE-) GENESIS RES & DEV CORP LTD.

Glenn M, Grigor MR, Molenaar AJ, Davis SR;

WPI; 2003-275306/27.  
P-PSDB; ABG75792.

New histatin polynucleotides and polypeptides expressed in bovine PT tissues, useful for treating bacterial or fungal infections in mammals, e.g. infections of the oral cavity, vagina, urethra, ear or skin, or PT systemic infections

Claim 1; Page 11; 17pp; English.

The invention discloses isolated histatin polynucleotides and polypeptides, which comprise sequences expressed in bovine. Histatins are histidine rich and specific to the salivary secretions. They are believed to function as part of the non-immune defence system, particularly in the oral cavity and have promise as therapeutic agents in humans with oral candidosis. Also disclosed are compositions comprising the histatin polypeptide or polynucleotide, and at least one component consisting of physiological or pharmaceutical carriers or immunostimulants. The histatin polynucleotide, polypeptide, or the cosmetic composition comprising the polypeptide, is useful for treating (e.g. gene therapy) a disorder in a mammal, particularly microbial or fungal infections. The histatin polynucleotide is also useful in genome mapping, physical mapping or in the positional cloning of genes. Specifically, the polypeptide or polynucleotide is useful for treating fungal or bacterial infections of the oral cavity (e.g. dental caries, plaque or tartar), vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis caused by *Pseudomonas aeruginosa*), mucosa or eye, as well as for treating systemic infections such as systemic *Candida* infection. The polypeptide or polynucleotide is also useful for veterinary applications e.g. for treating mastitis. The sequence presented is a cDNA encoding a bovine histatin like polypeptide.

Sequence 585 BP; 215 A; 111 C; 85 G; 174 T; 0 other;

Query Match 74.2%; Score 448.2; DB 25; Length 585;  
Best Local Similarity 99.3%; Pred. No. 3.7e-87;  
Matches 450; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 138 CTGACCTCCACCAATATGAGATCTTTCTTTGCTTCAATTATGGCTCTCATCCTAC 197
DB 133 CAGGACTCCACCAATATGAGATCTTTCTTTGCTTCAATTATGGCTCTCATCCTAC 192
QY 198 CATGATTAGAGCTGATTCATCTGAAGAGAAAGCTCAAGAGAAACGAAAAAATCATATG 257
DB 193 CATGATTAGAGCTGATTCATCTGAAGAGAAAGCTCAAGAGAAACGAAAAAATCATATG 252
QY 258 AGGATATTTTCAACATATACCATATCAAGATATCCATTAATTTATCTCTGCGTA 317
DB 253 AGGATATTTTCAACATATACCATATCAAGATATCCATTAATTTATCTCTGCGTA 312
QY 318 TCCATTTCTTAAATGCTGCTTAGTAATACAGGACATGATTAAGAGATTTTTCACA 377
DB 313 TCCATTTCTTAAATGCTGCTTAGTAATACAGGACATGATTAAGAGATTTTTCACA 372
QY 378 TCAATTTTCTTAAATGCTGCTTAGTAATACAGGACATGATTAAGAGATTTTTCACA 437

```

[illegible]

Query	Subject	Score	Length	Ident	Mismatch	Indel	Gap
138	CTGACATCCCAACCAATATGAAAGACTTTATCTTGTCTTCATATAGGCTCATCCATG	197	525	70.1%	93.4%	0	0
27	CAGAGCTCCCAAAATATGAAAGACTTTATCTTGTCTTCATATAGGCTCATCCATG	86	525	70.1%	93.4%	0	0
198	CATGATTAAGAGCTGATTCATCTGAAAGAAAGCTCACAGAAAGCGAAAGAAACATCAT	255	525	70.1%	93.4%	0	0
87	CATGATTAAGAGCTGATTCATCTGAAAGAAAGCTCACAGAAAGCGAAAGAAACATCAT	146	525	70.1%	93.4%	0	0
256	-----AGAGATATTTTCAACAATATCCAGCCATAT	285	525	70.1%	93.4%	0	0
147	TGATAGCTCCCAAGATTTCTTACTATATACAGAGAGATTTTCAACAATATCCAGCCATAT	206	525	70.1%	93.4%	0	0
286	CAAGCATATCCATTAATATATCTCTCCGCTATCATTTCTTTAAAGCGCTTAAGTA	345	525	70.1%	93.4%	0	0
207	CAAGCATATCCATTAATATATCTCTCCGCTATCATTTCTTTAAAGCGCTTAAGTA	266	525	70.1%	93.4%	0	0
346	CTACAGAGCATGATTAAGAGAGATTTTTCACATGATTTTCTCTCTCTCTCTCTCTCT	405	525	70.1%	93.4%	0	0
267	CTACAGAGCATGATTAAGAGAGATTTTTCACATGATTTTCTCTCTCTCTCTCTCTCT	326	525	70.1%	93.4%	0	0
406	GAAGACCATCTTTCAATGAAATGAAACCAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	465	525	70.1%	93.4%	0	0
327	GAAGACCATCTTTCAATGAAATGAAACCAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	386	525	70.1%	93.4%	0	0
466	CACATCTCTGGAATCAATATATCATATTTTAAACATATATATATATATATATATATAT	525	525	70.1%	93.4%	0	0
387	CACATCTCTGGAATCAATATATCATATTTTAAACATATATATATATATATATATATAT	446	525	70.1%	93.4%	0	0
526	GTAATTTGGTTCTCACTTTCTTTCTCTGTCACCTTACCATGATGCTTAATATATATAT	585	525	70.1%	93.4%	0	0
447	GTAATTTGGTTCTCACTTTCTTTCTCTGTCACCTTACCATGATGCTTAATATATATAT	506	525	70.1%	93.4%	0	0
586	ATCAAGCATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	604	525	70.1%	93.4%	0	0
507	ATCAAGCATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	525	525	70.1%	93.4%	0	0



QY 138 CTGGACTCCACCAATATGAGATCTTATCTTCTTATTAAGCTCTCATCTAGC 197  
 DB 56 CAGACCTCCACCAATATGAGATCTTATCTTCTTATTAAGCTCTCATCTAGC 115  
 QY 198 CATGATTAGAGCTGATTCATCTGTAAGAGAAAGCTCAGAGAAACGAAAAACATCATAG 257  
 DB 116 CATGATTAGAGCTGATTCATCTGTAAGAGAAAGCTCAGAGAAACGAAAAACATCATGT 175  
 QY 258 AGGATATTTTC 268  
 DB 176 ATGTATTCCTC 186

RESULT 7  
 ABX11392  
 ID ABX11392 standard; cDNA; 96 BP.  
 AC ABX11392;  
 DT 01-MAY-2003 (first entry)  
 XX  
 DE Bovine histatin like polypeptide, #5, cDNA.  
 XX  
 KW .Cow; gene; ss; histatin; bovine; salivary secretion; oral cavity;  
 KW non-immune defence system; oral candidosis; gene therapy;  
 KW microbial infection; fungal infection; dental caries; plaque; tartar;  
 KW cystic fibrosis; systemic infection; Candida infection; mastitis;  
 KW fungicide; antibacterial.  
 XX  
 OS Bos taurus.  
 XX  
 FH Key 1.96 Location/Qualifiers  
 FT CDS /\*tag= a  
 FT /product= "Histatin like polypeptide"  
 FT /partial  
 FT /note= "No start or stop codon shown"  
 XX  
 PN US2002164625-A1.  
 XX  
 PD 07-NOV-2002.  
 XX  
 PF 19-FEB-2002; 2002US-0079754.  
 XX  
 PR 23-AUG-1999; 99US-150330P.  
 PR 29-OCT-1999; 99US-162701P.  
 PR 22-AUG-2000; 2000US-0644190.  
 PR 27-OCT-2000; 2000US-0699146.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 XX  
 PI Glenn M, Grigor MR, Molenaar AJ, Davis SR;  
 XX  
 DR WPI; 2003-275306/27.  
 DR P-PSDB; ABG75795.  
 XX  
 PT New histatin polynucleotides and polypeptides expressed in bovine  
 PT tissues, useful for treating bacterial or fungal infections in mammals,  
 PT e.g. infections of the oral cavity, vagina, urethra, ear or skin, or  
 PT systemic infections  
 XX  
 PS Claim 1; Page 12; 17pp; English.  
 XX  
 CC The invention discloses isolated histatin polynucleotides and  
 CC polypeptides, which comprise sequences expressed in bovine. Histatins are  
 CC histidine rich and specific to the salivary secretions. They are believed  
 CC to function as part of the non-immune defence system, particularly in the  
 CC oral cavity and have promise as therapeutic agents in humans with oral  
 CC candidosis. Also disclosed are compositions comprising the histatin  
 CC polypeptide or polynucleotide, and at least one component consisting of  
 CC physiological or pharmaceutical carriers or immunostimulants. The  
 CC histatin polynucleotide, polypeptide, or the cosmetic composition

CC comprising the polypeptide, is useful for treating (e.g. gene therapy) a  
 CC disorder in a mammal, particularly microbial or fungal infections. The  
 CC histatin polynucleotide is also useful in genome mapping, physical  
 CC mapping or in the positional cloning of genes. Specifically, the  
 CC polypeptide or polynucleotide is useful for treating fungal or bacterial  
 CC infections of the oral cavity (e.g. dental caries, plaque or tartar),  
 CC vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis  
 CC caused by *Pseudomonas aeruginosa*), mucosa or eye, as well as for treating  
 CC systemic infections such as systemic Candida infection. The polypeptide  
 CC or polynucleotide is also useful for veterinary applications e.g. for  
 CC treating mastitis. The sequence presented is a cDNA encoding a bovine  
 CC histatin like polypeptide.  
 CC  
 SQ Sequence 96 BP; 33 A; 25 C; 10 G; 28 T; 0 other;  
 QY Query Match 14.4%; Score 87; DB 25; Length 96;  
 DB Best Local Similarity 94.7%; Pred. No. 1.1e-09;  
 DB Matches 90; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 265 TTTCACCAATATGAGATCTTATCTTCTTATTAAGCTCTCATCTAGC 324  
 DB 1 TTTCACCAATATGAGATCTTATCTTCTTATTAAGCTCTCATCTAGC 60  
 QY 325 CCTTAAATGCTGCTTACTACTACGACATGAT 359  
 DB 61 TCTTAAATGCTGCTTACTACTACGACATGAT 95

RESULT 8  
 AAV89526  
 ID AAV89526 standard; cDNA; 357 BP.  
 AC AAV89526;  
 XX  
 DT 15-FEB-1999 (first entry)  
 XX  
 DE EST clone CP294.  
 XX  
 KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
 KW gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9845436-A2.  
 XX  
 PD 15-OCT-1998.  
 XX  
 PF 10-APR-1998; 98WO-US06955.  
 XX  
 PR 10-APR-1997; 97US-0838821.  
 XX  
 PA (GENE) GENETICS INST INC.  
 XX  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 PI Racie LA, Spaulding V, Treacy M;  
 XX  
 DR WPI; 1999-070077/06.  
 XX  
 PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries.  
 XX  
 PS Claim 1; Page 245; 618pp; English.  
 XX  
 CC The present sequence represents a human expressed sequence tag (EST).  
 CC The polynucleotide, which is a secreted EST, and the encoded protein  
 CC are predicted to have useful biological activities which would make  
 CC them suitable for treating, preventing or ameliorating medical  
 CC conditions in humans and animals, although no supporting data is  
 CC given. Suggested activities include nutritional activity, immune  
 CC stimulating or suppressing activity, haematopoiesis regulating

CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
XX  
SQ Sequence 357 BP; 101 A; 74 C; 65 G; 117 T; 0 other;

Query Match 12.5%; Score 75.4; DB 20; Length 357;  
Best Local Similarity 59.4%; Pred. No. 4.3e-07;  
Matches 149; Conservative 0; Mismatches 96; Indels 6; Gaps 1;

QY 120 GAATTCATCTTTCATGACTGATCCACCAAAATGAGATCTTATCTTGTCTCAT 179  
DB |||||  
DB 48 GACTCTCTCTTGAGTAAAGAGCTACGCACTATGAAATTTTGTCTTGTCTTATG 107  
QY 180 TATGCTCTCATCCATGACCATATTAAGCTATTCATCTGAAGAAACGTACAGAA 239  
DB |||||  
DB 108 CTGGCTCTCATGATTTCCATGATTAAGCGTATTCACATGAAAGACATCATGGGTA 167  
QY 240 ACGGAAAAACATCATAGAGATATTTTCAACATACCCCATATCAACGATCCACT 299  
DB |||||  
DB 168 TAGAAGAAAATTCAT-----GAAAGCATCATTCACATCGAATTTCCATTTATG 221  
QY 300 AATTAATCTCTCGGTATCATTTCTTAAATGCTGTTAGTACTACAGACATGAT 359  
DB |||||  
DB 222 GGACTGTGATCAAAATTAATCTATGACATTAATGATCTTATGTAATATGCGCATGAT 281  
QY 360 TAGAGAGATTT 370  
DB |||||  
DB 282 TATAGAGGTTT 292

## RESULT 9

ACC51062  
ID ACC51062 standard; cDNA; 438 BP.

AC ACC51062;

DT 13-JUN-2003 (first entry)

DE Human Chimera coding sequence.

XX Human; GENSET; therapeutic; therapy; gene; ss.

OS Homo sapiens.

XX WO200294864-A2.

PD 28-NOV-2002.

PF 06-AUG-2001; 2001WO-IB01715.

PR 25-MAY-2001; 2001US-293574P.

PR 15-JUN-2001; 2001US-298698P.

PR 29-JUN-2001; 2001US-302277P.

PR 13-JUL-2001; 2001US-305456P.

XX (GENSET ) GENSET.

XX Benjamin S, Tanaka H;

XX WPI: 2003-129412/12.

XX P-PSDB; ABR48455.

PT New GENSET polynucleotides and polypeptides, useful for preparing a  
PT composition for treating GENSET-related disorders and as reagents in  
PT assays to quantitatively determined levels of GENSET expression in  
PT biological samples -  
PS Claim 1; Page 404-405; 505pp; English.

CC The present invention relates to novel human GENSET coding sequences

CC (ACC51060-ACC51115) and proteins (ABR48453-ABR48508). The GENSET  
CC sequences are useful for preparing a composition for treating  
CC GENSET-related disorders. They can also be used as markers for tissues in  
CC which the corresponding protein is preferentially expressed, as molecular  
CC weight markers on Southern gels, as chromosome markers or tags to  
CC identify chromosomes, and as reagents in assays to quantitatively  
CC determined levels of GENSET expression in biological samples.  
XX  
SQ Sequence 438 BP; 155 A; 80 C; 60 G; 143 T; 0 other;

Query Match 11.7%; Score 70.4; DB 25; Length 438;  
Best Local Similarity 69.9%; Pred. No. 5.4e-06;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACTGATCCACCAATATGAGATCTTATCTTGTCTCAT 179  
DB |||||  
DB 50 GACTCTCTCTTGAGTAAAGAGCTACGCACTATGAAATTTTGTCTTGTCTTATG 109  
QY 180 TATGCTCTCATCCATGACCATATTAAGCTATTCATCTGAAGAAACGTACAGAA 239  
DB |||||  
DB 110 CTGGCTCTCATGATTTCCATGATTAAGCGTATTCACATGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB |||||  
DB 170 TAGAAGAAAATTCAT 185

## RESULT 10

ABZ36425  
ID ABZ36425 standard; cDNA; 438 BP.

AC ABZ36425;

DT 21-FEB-2003 (first entry)

DE Human GENSET coding sequence SEQ ID 22.

XX Gastrointestinal; gene therapy; GENSET; heavy metal toxicity; cancer;

XX inflammatory disease; immune disorder; neuromuscular toxicity;

XX central nervous system; cardiovascular; gastrointestinal; gene; ss.

OS Homo sapiens.

XX WO200283938-A1.

PD 24-OCT-2002.

PF 18-APR-2001; 2001WO-IB00914.

PR 18-APR-2001; 2001WO-IB00914.

XX (GENSET ) GENSET.

XX Benjamin S, Tanaka H, Dumas Milne Edwards J, Robert S, Giordano J;

XX WPI: 2003-075548/07.

PT New GENSET polynucleotides and polypeptides, useful for treating heavy  
PT metal toxicity, cancer, inflammatory diseases, immune disorders, and  
PT the neuromuscular, CNS, cardiovascular or gastrointestinal effects of  
PT the toxicity -  
PS Claim 12; Page 295; 735pp; English.

CC The present invention relates to novel GENSET polynucleotides  
CC (ABZ36404-ABZ36911) encoding polypeptides (ABP75963-ABP76368). The  
CC polynucleotides and polypeptides are useful in screening and diagnostic  
CC assays for abnormal GENSET expression and/or biological activity. They  
CC are also useful for screening of compounds for treating or preventing  
CC GENSET-related disorders, such as heavy metal toxicity, cancer,  
CC inflammatory diseases, immune disorders, and the neuromuscular, central  
CC nervous system (CNS), cardiovascular or gastrointestinal effects of the

CC toxicity. The polynucleotides are useful for constructing or expanding  
CC chromosome maps.

XX Sequence 438 BP; 153 A; 80 C; 60 G; 143 T; 2 other;

SO Query Match 11.7%; Score 70.4; DB 25; Length 438;

Best Local Similarity 69.9%; Pred. No. 5.4e-06;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTCATGACGAGCTCCACCAATATGAAGCTTTATCTTGTCTTCAT 179

DB 50 GACTCTCTCTTGAGTAAAGAGCTCAGCCACTATGAAGTTTGTCTTGTCTTACT 109

QY 180 TATGCTCTCATCTTCAGCTAGATTAAGCTGATTCATCTGAAGAAAGCTCAGAGAA 239

DB 110 CTGGCTCTCATGATTTCCATGATTAAGCGCTGATTCACATGAAGAAAGACATCATGGTA 169

QY 240 ACGGAAAAAATCATCAT 255

DB 170 TAGAAGAAAAATTCAT 185

RESULT 11.

ABZ36590 ID ABZ36590 standard; cDNA; 438 BP.

XX ABZ36590;

DT 21-FEB-2003 (first entry)

DE Human GENSET coding sequence SEQ ID 358.

XX Cytostatic; antiinflammatory; nootropic; neuroprotective; cardiac;

KW gastrointestinal; gene therapy; GENSET; heavy metal toxicity; cancer;

KW inflammatory disease; immune disorder; neuromuscular; toxicity;

KW central nervous system; cardiovascular; gastrointestinal; gene; ss.

XX Homo sapiens.

OS WO200283898-A1.

PN 24-OCT-2002.

PF 18-APR-2001; 2001WO-IB00914.

PR 18-APR-2001; 2001WO-IB00914.

PI Bejanin S, Tanaka H, Dumas Milne Edwards J, Jobert S, Giordano J;

DR WPI; 2003-075548/07.

XX New GENSET polynucleotides and polypeptides, useful for treating heavy

PT metal toxicity, cancer, inflammatory diseases, immune disorders, and

PT the neuromuscular, CNS, cardiovascular or gastrointestinal effects of

PT the toxicity

PS Claim 12; Page 478; 735pp; English.

XX The present invention relates to novel GENSET polynucleotides

CC (ABZ34404-ABZ36911) encoding polypeptides (ABP75963-ABP76368). The

CC polynucleotides and polypeptides are useful in screening and diagnostic

CC assays for abnormal GENSET expression and/or biological activity. They

CC are also useful for screening of compounds for treating or preventing

CC GENSET-related disorders, such as heavy metal toxicity, cancer,

CC inflammatory diseases, immune disorders, and the neuromuscular, central

CC nervous system (CNS), cardiovascular or gastrointestinal effects of the

CC toxicity. The polynucleotides are useful for constructing or expanding

CC chromosome maps.

Query Match 11.7%; Score 70.4; DB 25; Length 438;  
Best Local Similarity 69.9%; Pred. No. 5.4e-06;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTCATGACGAGCTCCACCAATATGAAGCTTTATCTTGTCTTCAT 179

DB 50 GACTCTCTCTTGAGTAAAGAGCTCAGCCACTATGAAGTTTGTCTTGTCTTACT 109

QY 180 TATGCTCTCATCTTCAGCTAGATTAAGCTGATTCATCTGAAGAAAGCTCAGAGAA 239

DB 110 CTGGCTCTCATGATTTCCATGATTAAGCGCTGATTCACATGAAGAAAGACATCATGGTA 169

QY 240 ACGGAAAAAATCATCAT 255

DB 170 TAGAAGAAAAATTCAT 185

RESULT 12

AAH98658 ID AAH98658 standard; cDNA; 857 BP.

XX AAH98658;

DT 12-OCT-2001 (first entry)

DE Human EST-derived coding sequence SEQ ID NO: 515.

XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

KW diagnostics; forensic test; gene mapping; genetic disorder;

KW biodiversity; gene therapy; nutrition; ss.

XX Homo sapiens.

OS WO200154477-A2.

PN 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US02687.

PR 25-JUL-2000; 2000US-0491404.

PR 17-JUL-2000; 2000US-0617746.

PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.

PI (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

PI Cao Y, Drmanac RA, Zhang J, Werhman T;

DR WPI; 2001-476164/51.

DR P-PSDB; AAM23999.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising

PT antibodies and research use -

PS Claim 1; Page 537; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel

CC proteins from a variety of organisms, including human, dog, cat, horse,

CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea

CC urchin and tomato. These were derived from expressed sequence tags (ESTs)

CC from the organism of interest. They can be used in diagnostics,

CC forensics, gene mapping, identification of mutations, to assess

CC biodiversity and for nutritional purposes. The present sequence is a cDNA

CC of the invention.

SO Sequence 857 BP; 231 A; 194 C; 169 G; 263 T; 0 other;

Query Match 11.3%; Score 68.2; DB 22; Length 857;

Best Local Similarity 65.4%; Pred. No. 1.8e-05;  
Matches 100; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACTGACGCCAACCAATATGAGATCTTTATCTTCTTCAAT 179  
 Db 361 GATTCCTCTTTCGATTAAGAGCTCAGCCACTATGAGTTTGTGTTTGTCTTAAAT 420  
 QY 180 TATGCTCTCATCCAGCAGCATGATAGAGCTGATTCATCTGAGAGAAACGTCACAGAA 239  
 Db 421 CTGGCTCTCATGCTTTTCATGATGAGAGCTGATTCATCTGAGAGACATCATGGGTA 480  
 QY 240 ACGGAAAAACATCATAGAGATATTTTCAACA 272  
 Db 481 TAAAGAAAATTCATGAAAAGCATCATTCACA 513

## RESULT 13

AAAA8964  
 ID AAAA8964 standard; DNA; 552 BP.

XX AC AAAA8964;

XX DT 06-OCT-2000 (first entry)

XX DE Human statherin DNA.

XX KM Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;  
 KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;  
 KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;  
 KW cancer; leukemia; adenocarcinoma; melanoma; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 73..261  
 FT /\*tag= a  
 FT /product= Statherin

XX PN WO200024779-A1.

XX PD 04-MAY-2000.

XX PF 22-OCT-1999; 99WO-US24046.

XX PR 23-OCT-1998; 98US-0155209.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Tang YF, Corley NC, Guegler KJ, Patterson C;

XX DR WPI: 2000-350699/30.

XX DR P-PSDB; AAY94527.

XX PT Purified polypeptide used for treating or preventing a disorder  
 PT characterized by expression or activity of lysine-rich statherin  
 PT proteins -

XX PS Disclosure; Page 71; 75pp; English.

XX CC The present invention relates to human lysine-rich statherin protein  
 CC (LRSP) (AAY94526). The cDNA sequence encoding this protein was identified  
 CC through analysis of a cDNA library of breast tumour tissue (BRSTN0714).  
 CC The LRSP sequence was found to have homology with the DNA of human  
 CC statherin (the present sequence) and human basic histidine-rich protein  
 CC (AAY94528). Human statherin is a phosphoprotein that acts as an  
 CC inhibitor of precipitation of calcium phosphate salts in the oral cavity.  
 CC The LRSP polypeptide and its antagonists may be useful for treating or  
 CC preventing disorders associated with the activity of LRSP. Such  
 CC disorders include autoimmune/inflammatory disorders (for example AIDS,  
 CC allergies, asthma, diabetes mellitus), bacterial and fungal infection  
 CC and cancers (such as leukemia, adenocarcinoma, melanoma). Antibodies to  
 CC LRSP may be useful for diagnosis of the above disorders.

XX SQ Sequence 552 BP; 172 A; 122 C; 78 G; 180 T; 0 other;

Query Match 11.2%; Score 67.6; DB 21; Length 552;

Best Local Similarity 61.2%; Pred. No. 2.2e-05;  
 Matches 147; Conservative 0; Mismatches 84; Indels 9; Gaps 2;

QY 140 GGATCCCAACCAATATGAGATCTTTATCTTGTCTTCATATAGGCTCATCCAGCA 199  
 Db 59 GAACCAAGCCCACTATGAGATCTTCTTGTGCTTCATCTTGGCTCATAGGTTTCA 118  
 QY 200 TGATTAGAGCTGATTCATCTGAGAGAAACGTCACAGAAACGAAAAACATCATAGAG 259  
 Db 119 TGATTGAGCTGATTCATCTGAGAGAAATTTTGGCGTGAATGGAATTCGGTTAAG 178  
 QY 260 GATA-----TTTCAACATATACAGCCATATCAACGATATCACTAATATATCTGCTG 313  
 Db 179 GGTATGCGCCCTTATCAGCCAGTTCAGAAACCAACCACTATACCCACATATACCAACAC 238  
 QY 314 CGT---ATCCATTTTCCTTAAATGCTGTTAGTATCATACAGACATGATGAGAGATT 370  
 Db 239 AATACCAACAAATATATCCTTTTAAATATCATCATGTAACGACAGACATGATTAATGAGCTT 298

## RESULT 14

AAV89600  
 ID AAV89600 standard; cDNA; 203 BP.

XX AC AAV89600;

XX DT 15-FEB-1999 (first entry)

XX DE EST clone CP41.

XX KM Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
 KW tissue growth; activin; inhibin; chemokinesis; chemokinesis; haemostatic;  
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
 KW gene therapy; ss.

XX OS Homo sapiens.

XX PN WO9845436-A2.

XX PD 15-OCT-1998.

XX PF 10-APR-1998; 98WO-US06955.

XX PR 10-APR-1997; 97US-0838821.

XX PA (GENY) GENETICS INST INC.

XX PI Agostino MJ, Jacobs K, Lavalie ER, McCoy JM, Merberg D;

XX PI Racie LA, Spaulding V, Treacy M;

XX DR WPI: 1999-070077/06.

XX PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries.

XX PS Claim 1; Page 269-270; 618pp; English.

XX CC The present sequence represents a human expressed sequence tag (EST).  
 CC The polynucleotide, which is a secreted EST, and the encoded protein  
 CC are predicted to have useful biological activities which would make  
 CC them suitable for treating, preventing or ameliorating medical  
 CC conditions in humans and animals, although no supporting data is  
 CC given. Suggested activities include nutritional activity, immune  
 CC stimulating or suppressing activity, haematopoiesis regulating  
 CC activity, tissue growth activity, activin/inhibin activity,  
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
 CC activity, receptor/ligand activity, anti-inflammatory activity,  
 CC cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The polynucleotide may also be useful for gene therapy.

XX SQ Sequence 203 BP; 59 A; 43 C; 39 G; 62 T; 0 other;

Query Match 11.1%; Score 67.2; DB 20; Length 203;  
Best Local Similarity 68.4%; Pred. No. 2.3e-05;  
Matches 93; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 120 GAATTTATCTTTGATGAGTCCAGCCAAATATGAATCTTTATCTTTGCTTCAT 179  
DB 47 GATTTCTCTTGTAGTAAAGAGTCCAGCCAACTATTAAGTTTCTTTTGTCTTAT 106  
QY 180 TATGCTCTCATCTGACATGATTAGAGCTGATTCATGTAAGAGAAAGTACAGAA 239  
DB 107 CTGGCTCTCATGCTTTCCATGACTGAGCTGATTCATGCAAGAGATCATGAGTA 166  
QY 240 ACGGAAAAAATCATCAT 255  
DB 167 TAAAGAAAAATTCAT 182

RESULT 15  
AAA48965  
ID AAA48965 standard; DNA; 491 BP.  
XX  
AC AAA48965;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human basic histidine-rich protein DNA.  
XX  
KM lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;  
KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;  
KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;  
KW cancer; leukemia; adenocarcinoma; melanoma; de.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 38..193  
FT /\*tag= a  
FT /product= Basic\_histidine-rich\_protein  
XX  
PN W0200024779-A1.  
XX  
PD 04-MAY-2000.  
XX  
PF 22-OCT-1999; 99WO-US24046.  
XX  
PR 23-OCT-1998; 98US-0155209.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Tang YT, Corley NC, Guegler KJ, Patterson C;  
XX  
PI WPI; 2000-350699/30.  
XX  
DR P-PSDB; AA94528.  
XX  
XX  
PT Purified polypeptide used for treating or preventing a disorder  
PT characterized by expression or activity of lysine-rich statherin  
PT proteins -  
XX  
PS Disclosure; Page 72; 75pp; English.  
XX  
XX The present invention relates to human lysine-rich statherin protein  
XX (LRSP) (AA94526). The cDNA sequence encoding this protein was identified  
XX through analysis of a cDNA library of breast tumour tissue (BRSTNOT14).  
XX The LRSP sequence was found to have homology with human statherin  
XX protein (AA194527) and human basic histidine-rich protein (the present  
XX sequence). Human statherin is a phosphoprotein that acts as an inhibitor  
XX of precipitation of calcium phosphate salts in the oral cavity. The LRSP  
XX polypeptide and its antagonists may be useful for treating or preventing  
XX disorders associated with the activity of LRSP. Such disorders include  
XX autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,  
XX diabetes mellitus), bacterial and fungal infection and cancers (such as  
XX leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be useful  
XX for diagnosis of the above disorders.

XX  
SQ Sequence 491 BP; 151 A; 90 C; 87 G; 163 T; 0 other;  
Query Match 10.4%; Score 62.6; DB 21; Length 491;  
Best Local Similarity 66.9%; Pred. No. 0.00026;  
Matches 89; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 140 GGAATCCACCAATATGAAGATCTTTATCTTTGCTTCATTATGAGCTCTCATGCCA 199  
DB 24 GAATTCACCACTATGAAGTTTGTGCTTTAATCTTGCTCATGCTTCCA 83  
QY 200 TGAATTAGGCTGATTCATCTGAAAGAAAGTCAAGAAACGAAACATCATAGAG 259  
DB 84 TGAATGAGCTGATTCATGCAAGAGATCATGAGTAAAGAAAAATTCATGAAA 143  
QY 260 GATATTTTCACA 272  
DB 144 AGCATCATTCACA 156

Search completed: December 13, 2003, 01:45:50  
Job time : 254 secs





QY	Db	QY	Db
401	GTGTTGAAAACCATCTTTCAATGATATAAACAAAGAAAAAAAATTCGTCAGATGAGTGG	460	1404
1404	CTTTATTTATATATTTATTAATTATATACAGCTAGATTAATTTACTTTAAGAACTA	1463	
461	CACAAACATCTGGAGATCAAAATTCATAATTTTAAACATTAATTAAGATAGT	514	
1464	TAAATAATAGCTCTAGAGATTAATTAAGCTCTATATAAATAATTTATTAAGATTAATT	1517	

RESULT 2  
US-09-367-895-29/c

```

TITLE OF INVENTION: ANTISENSE INTRON INHIBITION OF STARCH BRANCHING ENZYME
FILE REFERENCE: 076883/0112
CURRENT APPLICATION NUMBER: US/09/367,895
CURRENT FILING DATE: 1999-12-08
PRIOR APPLICATION NUMBER: PCT/IB98/00270
PRIOR FILING DATE: 1998-02-23
PRIOR APPLICATION NUMBER: GB/9703663.6
PRIOR FILING DATE: 1997-02-21
PRIOR APPLICATION NUMBER: GB/9706060.2
PRIOR FILING DATE: 1997-03-24
NUMBER OF SEQ ID NOS: 43
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 29
LENGTH: 11469
TYPE: DNA
ORGANISM: Solanum tuberosum
FEATURE:
NAME/KEY: CDS
LOCATION: (2132..2209, 3375..3494, 3812..4033, 4538..4606,
LOCATION: 4753..5022, 5241..6146, 6345..6461, 6670..6732,
LOCATION: 7026..7133, 7510..7611, 7784..7852, 7998..8078,
LOCATION: 8321..8437, 9235..9630)
US-09-367-895-29

```

[illegible]

RESULT 3  
 US-08-981-803-29/c  
 Sequence 29, Application US/08981803  
 Patent No. 6147279  
 GENERAL INFORMATION:  
 APPLICANT: POLISEN, PETER  
 TITLE OF INVENTION: INHIBITION OF GENE EXPRESSION  
 FILE REFERENCE: 674509-2004  
 CURRENT APPLICATION NUMBER: US/08/981,803  
 CURRENT FILING DATE: 1997-04-17  
 EARLIER APPLICATION NUMBER: PCT/EP96/03052

```

:
: EARLIER FILING DATE: 1996-07-12
: EARLIER APPLICATION NUMBER: 9514435.8
: EARLIER FILING DATE: 1995-07-14
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 29
:
: LENGTH: 11478
:
: TYPE: DNA
: ORGANISM: Solanum tuberosum
:
: US-08-961-803--29

```

Query Match	Score	DB	Length
6.8%	41	3	11478

	Matches	95;	Conservative	0;	Mismatches	90;	Indels	0;	Gaps	0;
Qy	368	TTTTTCA	CAATGATTTTTCCTACTCTTTCTGTGTGTGAAAA	CCATCTTTCA	AATGAT	427				
Db	11381	TTTGACATCTTTTCTGTATATAATTTCTTCTTTCA	TATACAACCAACCAATGAAG	11322						
Qy	428	AAAA	CAAGAAAAAAATCAGTCAAGTAGTGCACACACATACTTGGANATCA	AATATC	487					
Db	11321	AAAA	CAATGAAGAAATCATATAGAAAAGGAAAGATTTTCACTCATTTGA	AGCGAA	11264					
Qy	488	AATATTTT	AAAA	CATATAATGATCTCTGAACATGTAATGTGTTTCTACTTTCTTT	547					
Db	11261	AATCTACT	AGAAAGATGATGTGAACCTTGTGAAACCTTCATTA	TGTGACACCCCTCGTT	11202					
Qy	548	TCTCT	552							
Db	11201	CTACT	11197							

RESULT 4  
US-08-983-440-29/C

Patent No. 6232122  
 GENERAL INFORMATION:  
 APPLICANT: POLISSEN, Peter  
 TITLE OF INVENTION: INHIBITION OF GENE EXPRESSION  
 FILE REFERENCE: 674509-2003  
 CURRENT APPLICATION NUMBER: US/08/993,440  
 CURRENT FILING DATE: 1998-04-17  
 EARLIER APPLICATION NUMBER: 9514437.4  
 EARLIER FILING DATE: 1995-07-14  
 EARLIER APPLICATION NUMBER: PCT/EP96/03053  
 EARLIER FILING DATE: 1996-07-12  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 29  
 LENGTH: 11478  
 TYPE: DNA  
 ORGANISM: Solanum tuberosum  
 FEATURE:  
 NAME/KEY: variation  
 LOCATION: (1)..(11478)  
 OTHER INFORMATION: B stands for G or C or T/U  
 FEATURE3:  
 NAME/KEY: variation  
 LOCATION: (1)..(11478)  
 OTHER INFORMATION: R stands for G or A  
 FEATURE:  
 NAME/KEY: variation  
 LOCATION: (1)..(11478)  
 OTHER INFORMATION: W stands for A or T/U  
 FEATURE:  
 NAME/KEY: variation  
 LOCATION: (1)..(11478)  
 OTHER INFORMATION: M stands for A or C  
 OTHER INFORMATION: M stands for A or C  
 US-08-993-440-29

Query Match 6.8%; Score 41; DB 3; Length 11478;  
Best Local Similarity 51.4%; Pred. No. 1.1;  
Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 368 TTTTTCACATGATTTTTCCTCTCTCTGTTGTTGAAAACCTTTCAATGAAT 427  
DB 11381 TTTGACATCTTTTCTGTATTAATTTTCTTTTCATATACCAACCAATGAAG 11322

QY 428 AAAACAAAGAAAAAATCAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 487  
DB 11321 AAAACATGAGAAATTCATAGAAAGAAAGAAATTTTCATCTCATTTGAAACGAA 11262

QY 488 AATATTTTAAACATATATATATAGTCTCTGAACTATGATATGTTGTTCTACTT 547  
DB 11261 AATCTCTAGAGAGATGATGATGAACTCTGAACTTTTCATCTGACACCTCTGTT 11202

QY 548 TCTCT 552  
DB 11201 CTAAT 11197

RESULT 5  
US-08-973-462-1  
Sequence 1, Application US/08973462B  
Patent No. 6191270  
GENERAL INFORMATION:  
APPLICANT: DRUIHE, PIERRE  
APPLICANT: DAUBERSTES, PIERRE  
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
FILE REFERENCE: 0660-0125-0 PCT  
CURRENT APPLICATION NUMBER: US/08/973,462B  
CURRENT FILING DATE: 1998-02-06  
EARLIER APPLICATION NUMBER: PCT/FR96/00894  
EARLIER FILING DATE: 1996-06-12  
EARLIER APPLICATION NUMBER: FR 95/07007  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1  
LENGTH: 6152  
TYPE: DNA  
ORGANISM: P. falciparum  
US-08-973-462-1

Query Match 6.8%; Score 40.8; DB 3; Length 6152;  
Best Local Similarity 53.0%; Pred. No. 1;  
Matches 87; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 420 AAATGAATAAAAAAGAAAAAATCAGTACAGTACAGTACAGTACAGTACAGTAC 479  
DB 309 AAAAAAAGAAAAAATCAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 368

QY 480 CAAATATCAATTTTAAACATATATATGATGATGCTCTGAACTATGATGTTGTTCT 539  
DB 369 TAT 428

QY 540 CTTTCTCTCTCTGCTACCTACATGATGATGATGATGATGATGATGATGATGAT 583  
DB 429 TATATCTTTTCTGATATATATATATATATATATATATATATATATATATAT 472

RESULT 6  
5496550-9  
Patent No. 5496550  
APPLICANT: WALLACH, MICHAEL; PUGATSCH, THEA; MENCHER, DAVID  
TITLE OF INVENTION: METHOD OF REDUCING THE OUTPUT OF EIMERIA  
OOCYSTS FROM A MEMBORN CHICK  
NUMBER OF SEQUENCES: 10  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/108,763  
FILING DATE: 17-AUG-1993  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 642,219  
FILING DATE: 16-JAN-1991  
APPLICATION NUMBER: 310,603  
FILING DATE: 14-FEB-1989  
APPLICATION NUMBER: 155,245  
FILING DATE: 12-FEB-1988  
APPLICATION NUMBER: 896,611  
FILING DATE: 14-AUG-1986  
SEQ ID NO: 9  
LENGTH: 465  
5496550-9

Query Match 6.7%; Score 40.4; DB 6; Length 465;  
Best Local Similarity 51.7%; Pred. No. 0.64;  
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 425 AATTAACAAAGAAAAAATCAGTACAGTACAGTACAGTACAGTACAGTAC 484  
DB 149 AAGCAAGAAACACCTCTTATGACATGATGATGATGATGATGATGATGATGAT 208

QY 485 ATCAATATTTTAAACATATATATATAGTCTCTGAACTATGATGATGATGATGATG 544  
DB 209 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 268

QY 545 TTTTCTCTGCTACCTACATGATGATGATGATGATGATGATGATGATGATGATGAT 602  
DB 269 GTGCTTTTGTGATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 326

RESULT 7  
5496550-7  
Patent No. 5496550  
APPLICANT: WALLACH, MICHAEL; PUGATSCH, THEA; MENCHER, DAVID  
TITLE OF INVENTION: METHOD OF REDUCING THE OUTPUT OF EIMERIA  
OOCYSTS FROM A MEMBORN CHICK  
NUMBER OF SEQUENCES: 10  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/108,763  
FILING DATE: 17-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 642,219  
FILING DATE: 16-JAN-1991  
APPLICATION NUMBER: 310,603  
FILING DATE: 14-FEB-1989  
APPLICATION NUMBER: 155,245  
FILING DATE: 12-FEB-1988  
APPLICATION NUMBER: 896,611  
FILING DATE: 14-AUG-1986  
SEQ ID NO: 7  
LENGTH: 466  
5496550-7

Query Match 6.7%; Score 40.4; DB 6; Length 466;  
Best Local Similarity 51.7%; Pred. No. 0.64;  
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 425 AATTAACAAAGAAAAAATCAGTACAGTACAGTACAGTACAGTACAGTAC 484  
DB 150 AAGCAAGAAACACCTCTTATGACATGATGATGATGATGATGATGATGATGAT 209

QY 485 ATCAATATTTTAAACATATATATATAGTCTCTGAACTATGATGATGATGATGATG 544  
DB 210 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 269

QY 545 TTTTCTCTGCTACCTACATGATGATGATGATGATGATGATGATGATGATGATGAT 602  
DB 270 GTGCTTTTGTGATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 327

RESULT 8  
US-08-973-462-1/c  
Sequence 1, Application US/08973462B  
Patent No. 6191270

```

? GENERAL INFORMATION:
? APPLICANT: DROUHE, PIERRE
? APPLICANT: DAUBERTSIES, PIERRE
? TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
? FILE REFERENCE: 0660-0125-0 PCT
? CURRENT APPLICATION NUMBER: US/08/973,462B
? CURRENT FILING DATE: 1998-02-06
? EARLIER APPLICATION NUMBER: PCT/FR96/00894
? EARLIER FILING DATE: 1996-06-12
? EARLIER APPLICATION NUMBER: FR 95/07007
? EARLIER FILING DATE: 1995-06-13
? NUMBER OF SEQ ID NOS: 29
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 1
? LENGTH: 6152
? TYPE: DNA
? ORGANISM: P. falciparum
? IS-08-973-462-1

```

[illegible]

```

RESULT 9
US-09-702-705-214/C
; Sequence 214, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaicanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darriick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 214
; LENGTH: 789
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-702-705-214

```

Query Match	6.6%	Score 40;	DB 4;	Length 789;
Best Local Similarity	46.2%	Pred. No. 0.91;		
Matches 133;	Conservative 0;	Mismatches 155;	Indels 0;	Gaps 0

  

QY	263	ATTITCAACAATACCACCATATCAAGATATCCATATTTATCTCTCGGTATCCAT	322
Db	523	ATTTTCATTATTACCATCATATCTTCTTAATGAGAAATATATGATATGAACATCCAT	464
QY	323	TTTCCTTAAATAGCTGCTTAGTACTACTACGAGACATGATTAGAGAGATTTTTCACAATGATT	382

Db	463	GCACATAACCCCTTATCTCTTCATGAATGTTTCCTTAAGGTAACTTTCAAGTGA	404
Qy	383	TTTCTACTCTTTCCTGTTGTGTGAAAAACAATCTTTCAATGAATAAACAAAGAAAA	442
Db	403	ATTGCTATCTTCTCTGACTTTCACAAAGCTCTTTATCTTTCAAAAGTTTGTATATT	344
Qy	443	AAATCACTGAATTAATTGACAAACAATCACTGGAATCAATATCAATATTTTAAACAT	502
Db	343	GAAAAAGCAACCTCTTATTGAACACATAGTAGGTCGAATTTACACACATAGCAGAGA	284
Qy	503	AATTAATGATAGCTCTGCAACTATGAAATTTGGTTCTCTTCTTCTTCTTCT	550
Db	283	GGTATATATAGGTTGAACATTTCTTTGTCTATATTAAGACTTTGTATATCT	236

```

? RESULT 10
? US-09-736-457-214/c
? Sequence 214, Application US/09736457
? Patent No. 6509448
? GENERAL INFORMATION:
? APPLICANT: Wang, Tongtong
? APPLICANT: Bangur, Chaitanya S.
? APPLICANT: Lodes, Michael A.
? APPLICANT: Fanger, Gary
? APPLICANT: Vedvick, Tom
? APPLICANT: Carter, Darrick
? APPLICANT: Retter, Marc
? APPLICANT: Mannion, Jane
? APPLICANT: Fan, Liqun
? APPLICANT: Wang, Aijun
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
? TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
? FILE REFERENCE: 210121.478C15
? CURRENT APPLICATION NUMBER: US/09/736,457
? CURRENT FILING DATE: 2000-12-13
? NUMBER OF SEQ ID NOS: 1864
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 214
? LENGTH: 789
? TYPE: DNA
? ORGANISM: Homo sapien
? US-09-736-457-214

```

Query Match	6.6%	Score 40	DB 4	Length 789
Best Local Similarity	46.2%	Pred. No. 0.91	Mismatches 155	Indels 0
Matches 133	Conservative	0	Gaps 0	
Qy	263	ATTTCACAAATACAGCAGCATATCACCATATCCACTAAATATCTCTGCGATCCAT	3222	
Db	523	ATTTCATTATTAACATCATATCTTCATATGAAGAATATATGATATGAACATCCAT	4646	
Qy	323	TTTCCTAAATAGCTGCTTAGTAACACGACATGATTAGAGATTTTTCACATGATT	3822	
Db	463	GCACATTAACCCCTGTATCTCTCATGATGTTTCTTAGGGTAAGTTCTTAGAAGTGA	4064	
Qy	383	TTTCTACTCTTTCGTGTGTGTTGAAAAACAATCTTTCAATGAATAAACAAAGAAAAA	442	
Db	403	ATTGCTATCTTCTCTTGAATCTTACAAAGCTCTTTAATCTTTCAAAAGTTTGTATAT	344	
Qy	443	AAATCGTAAAGTAGTGGACAAACACATCTTGGATCAATCAATATTTTAAAAAT	502	
Db	343	GAAGAAAGACAACTCTTATTGAACACATAGAGGTGCAATTTACACACATAGCGACAGGA	284	
Qy	503	AAATAGATAGCTCTGAACTATGTAATGTGTTTCTACTCTTTCTT	550	
Db	283	GGTAATATAGGTAAACATTTCTTGTCTATATTAAGCTTTGTATCT	236	

RESULT 11  
US-09-417-485D-7/c  
; Sequence 7, Application US/09417485D  
; Patent No. 6541202  
; GENERAL INFORMATION:

APPLICANT: Long, David M.  
APPLICANT: Metz, Anneke M.  
APPLICANT: Love, Ruscelle A.  
TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes  
FILE REFERENCE: 47714-5009-US  
CURRENT APPLICATION NUMBER: US/09/417,485D  
CURRENT FILING DATE: 2002-06-14  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 7  
LENGTH: 4185  
TYPE: DNA  
ORGANISM: Plasmodium falciparum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(2382)  
OTHER INFORMATION: Partial TERT gene  
US-09-417-485D-7

Query Match 6.6%; Score 39.8; DB 4; Length 4185;  
Best Local Similarity 55.4%; Pred. No. 1.6;  
Matches 77; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 387 CTACTCTTCTGTTGTTGAAAAACCATCTTCAATGAATTAACAAAGAAAAAAT 446  
Db 2662 CCAATCTATGTTTGACACCAAGCCAGATTATTAATAAAGAAAAAATTAATAAATA 2603  
Qy 447 CAGTCAAGTAGTTCACAAACATCTTGAAATCAATATCTTTTAAACATATA 506  
Db 2602 TATGCAATCCAAATATATATTTTAAACAAATGCAATGCAAAAAAATAAATA 2543  
Qy 507 ATGATAGTCTCTGAACAT 525  
Db 2542 ATAATGAATGTATTAAT 2524

RESULT 12  
US-09-482-273-47/c  
Sequence 47, Application US/09482273  
Patent No. 6534631  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 71 Human Secreted Proteins  
FILE REFERENCE: P2030P1  
CURRENT APPLICATION NUMBER: US/09/482,273  
CURRENT FILING DATE: 2000-01-13  
EARLIER APPLICATION NUMBER: PCT/US99/15849  
EARLIER FILING DATE: 1999-07-14  
EARLIER APPLICATION NUMBER: 60/092,921  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/092,922  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/092,956  
EARLIER FILING DATE: 1998-07-15  
NUMBER OF SEQ ID NOS: 267  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 47  
LENGTH: 6107  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5749)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5892)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5896)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:

NAME/KEY: SITE  
LOCATION: (5906)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5957)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5966)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-482-273-47

Query Match 6.6%; Score 39.8; DB 4; Length 6107;  
Best Local Similarity 59.1%; Pred. No. 1.7;  
Matches 68; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 432 CAAGAAAAAAATTCAGTCAAGTAGTTCACACACATACCTGAAATCAATATCAATA 491  
Db 2389 CAAAAAAGAAAGTAAGTACTAGACACTGGACCTGGACATCCAAATG 2330  
Qy 492 TTTTAAACATTAATATGATAGTCTCTGACTATGTAATGCTTCTTACTTT 546  
Db 2329 ATATGTAGATTATCTGATTTTACTGCAATATGAGCATTAATCTTCAATTTT 2275

RESULT 13  
US-09-417-485D-5/c  
Sequence 5, Application US/09417485D  
Patent No. 6541202  
GENERAL INFORMATION:  
APPLICANT: Long, David M.  
APPLICANT: Metz, Anneke M.  
APPLICANT: Love, Ruscelle A.  
TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes  
FILE REFERENCE: 47714-5009-US  
CURRENT APPLICATION NUMBER: US/09/417,485D  
CURRENT FILING DATE: 2002-06-14  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 5  
LENGTH: 10640  
TYPE: DNA  
ORGANISM: Plasmodium falciparum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (834)..(7385)  
OTHER INFORMATION: TERT gene  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1821)..(1837)  
OTHER INFORMATION: m at position 1821 = a or c; w at position 1837 =  
OTHER INFORMATION: a or t. Xaa (amino acid) at position 330 = Leu or  
OTHER INFORMATION: Ile; Xaa at position 335 = Asp or Gly.  
US-09-417-485D-5

Query Match 6.6%; Score 39.8; DB 4; Length 10640;  
Best Local Similarity 55.4%; Pred. No. 2;  
Matches 77; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 387 CTACTCTTCTGTTGTTGAAAAACCATCTTCAATGAATTAACAAAGAAAAAAT 446  
Db 7664 CCAATCTATGTTTGACACCAAGCAAGATTATTAATAAAGAAAAAATAAATA 7605  
Qy 447 CAGTCAAGTAGTTCACAAACATCTTGAAATCAATATCTTTTAAACATATA 506  
Db 7604 TATGCAATCCAAATATATATTTTAAACAAATGCAAAAAAATAAATA 7545  
Qy 507 ATGATAGTCTCTGAACAT 525  
Db 7544 ATAATGAATGTATTAAT 7526



Db 1 GAAGTATTTTCAGTTCATAATAAGATCTC

QY 61 GGAATTCAGGATATTAAACAGCAGCTTTCTAGCAAGAACATCTCTGAAGCATCAG 120  
DB 61 GGAATTCAGGATATTAAACAGCAGCTTTCTAGCAAGAACATCTCTGAAGCATCAG 120  
QY 121 AATTTCATCTTTCATGCTGAGCTCCCAATATGAAGATCTTATCTTGTCTTAT 180  
DB 121 AATTTCATCTTTCATGCTGAGCTCCCAATATGAAGATCTTATCTTGTCTTAT 180  
QY 181 ATGGCTCTCATCTCAGCAGCATGATTAGAGCTGATTCATCTGAGAGAAAGCTCAGAGAA 240  
DB 181 ATGGCTCTCATCTCAGCAGCATGATTAGAGCTGATTCATCTGAGAGAAAGCTCAGAGAA 240  
QY 241 CGAGAAAAACATCATAGAGATATTTTCAACATACAGCCATATCAAGATATCCACTA 300  
DB 241 CGAGAAAAACATCATAGAGATATTTTCAACATACAGCCATATCAAGATATCCACTA 300  
QY 301 AATTATCTCTGCTGATTCATCTTCTTAAATGCTCTTAAATGCTCTTAAATGCTCTTAA 360  
DB 301 AATTATCTCTGCTGATTCATCTTCTTAAATGCTCTTAAATGCTCTTAAATGCTCTTAA 360  
QY 361 AGAGAGATTTTTCACATGATTTTCTCTACTCTTCTGTTGTTGAAAACATCTTCA 420  
DB 361 AGAGAGATTTTTCACATGATTTTCTCTACTCTTCTGTTGTTGAAAACATCTTCA 420  
QY 421 AATGAATAAACAAAGAAAAAAATGCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 480  
DB 421 AATGAATAAACAAAGAAAAAAATGCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 480  
QY 481 AATATATATATTTTAAACATATATATGATGCTGATGATGATGATGATGATGATGATG 540  
DB 481 AATATATATATTTTAAACATATATATGATGCTGATGATGATGATGATGATGATGATG 540  
QY 541 TTCTTTCTCTGCTGATTCATCTTCTTAAATGCTCTTAAATGCTCTTAAATGCTCTTAA 600  
DB 541 TTCTTTCTCTGCTGATTCATCTTCTTAAATGCTCTTAAATGCTCTTAAATGCTCTTAA 600  
QY 601 AAAA 604  
DB 601 AAAA 604

RESULT 2  
US-10-079-754A-1  
; Sequence 1, Application US/10079754A  
; Publication No. US20020164625A1  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Grigor, Murray R.  
; APPLICANT: Molenaar, Adrian J.  
; APPLICANT: Davis, Stephen R.  
; TITLE OF INVENTION: Compositions Isolated from Bovine  
; TITLE OF INVENTION: Mammary Gland and Methods for Their Use  
; FILE REFERENCE: 11000.1068  
; CURRENT APPLICATION NUMBER: US/10/079, 754A  
; CURRENT FILING DATE: 2002-02-19  
; PRIOR APPLICATION NUMBER: US 09/699,146  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US 60,162,701  
; PRIOR FILING DATE: 1999-10-29  
; PRIOR APPLICATION NUMBER: US 09/644,190  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: US 60,150,330  
; PRIOR FILING DATE: 1999-08-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 505  
; TYPE: DNA  
; ORGANISM: Bovine  
US-10-079-754A-1  
Query Match 77.1%; Score 465.4; DB 14; Length 505;  
Best local Similarity 99.8%; Pred. No. 6e-95;

Matches 466; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 138 CTGAGCTCCACCAATATGAGATCTTATCTTGTCTTCAATATGAGCTCTGATCAGC 197  
DB 28 CAGAGCTCCACCAATATGAGATCTTATCTTGTCTTCAATATGAGCTCTGATCAGC 87  
QY 198 CATGATTAGAGCTGATTCATCTGAGAGAAAGCTCAGAGAAAGCTCAGAGAAAGCTCAG 257  
DB 88 CATGATTAGAGCTGATTCATCTGAGAGAAAGCTCAGAGAAAGCTCAGAGAAAGCTCAG 147  
QY 258 AGATATTTTCAACATATACAGCATATGAGATATGAGATATGAGATATGAGATATGAG 317  
DB 148 AGATATTTTCAACATATACAGCATATGAGATATGAGATATGAGATATGAGATATGAG 207  
QY 318 TCATTTCTCTTAAATGCTGCTAGTATGATCAAGACATGATATGAGATATGAGATATGAG 377  
DB 208 TCATTTCTCTTAAATGCTGCTAGTATGATCAAGACATGATATGAGATATGAGATATGAG 267  
QY 378 TGAATTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 437  
DB 268 TGAATTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 327  
QY 438 AAAAAAATCAGCAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 497  
DB 328 AAAAAAATCAGCAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 387  
QY 498 AACATATATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 557  
DB 388 AACATATATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 447  
QY 558 TTACATGATGCTTAAATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 604  
DB 448 TTACATGATGCTTAAATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 494

RESULT 3  
US-10-079-754A-2  
; Sequence 2, Application US/10079754A  
; Publication No. US20020164625A1  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Grigor, Murray R.  
; APPLICANT: Molenaar, Adrian J.  
; APPLICANT: Davis, Stephen R.  
; TITLE OF INVENTION: Compositions Isolated from Bovine  
; TITLE OF INVENTION: Mammary Gland and Methods for Their Use  
; FILE REFERENCE: 11000.1068  
; CURRENT APPLICATION NUMBER: US/10/079, 754A  
; CURRENT FILING DATE: 2002-02-19  
; PRIOR APPLICATION NUMBER: US 09/699,146  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US 60,162,701  
; PRIOR FILING DATE: 1999-10-29  
; PRIOR APPLICATION NUMBER: US 09/644,190  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: US 60,150,330  
; PRIOR FILING DATE: 1999-08-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 585  
; TYPE: DNA  
; ORGANISM: Bovine  
US-10-079-754A-2

Query Match 74.2%; Score 448.2; DB 14; Length 585;  
Best local Similarity 99.3%; Pred. No. 4.7e-91;  
Matches 450; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 138 CTGAGCTCCACCAATATGAGATCTTATCTTGTCTTCAATATGAGCTCTGATCAGC 197  
DB 133 CAGAGCTCCACCAATATGAGATCTTATCTTGTCTTCAATATGAGCTCTGATCAGC 192



QY 198 CATGATTAGAGCTGATTCATCTTGAAGAGAAAGCTCAAGGAAACGGAACAAATCATCATAG 257  
|  
|  
|  
Db 193 CATGATTAGAGCTGATTCATCTGGAAGAGAAAGCTCAAGGAAACGGAACAAATCATCATAG 252  
|  
|  
|  
QY 258 AGGATATTTCGAACATACAGCAATATCAAGATATCCATTAATTCCTCCGCGCA 317  
|  
|  
|  
Db 253 AGGATATTTCGAACATACAGCAATATCAAGATATCCATTAATTCCTCCGCGCA 312  
|  
|  
|  
QY 318 TCCATTTTCCTTAAATGCTGCTTAGTAACTACAGACATGATTAGAGATTTTTCACAA 377  
|  
|  
|  
Db 313 TCCATTTTCCTTAAATGCTGCTTAGTAACTACAGACATGATTAGAGATTTTTCACAA 372  
|  
|  
|  
QY 378 TGATTTTTCCTTAAATGCTGCTTAGTAAATCAATCTTTCAATTAATTAACAAAGA 437  
|  
|  
|  
Db 373 TGATTTTTCCTTAAATGCTGCTTAGTAAATCAATCTTTCAATTAATTAACAAAGA 432  
|  
|  
|  
QY 438 AAAAAAATCACTAGTACAGTACAGCAATCTTGAATCAATCAATATTTTNA 497  
|  
|  
|  
Db 433 AAAAAAATCACTAGTACAGTACAGCAATCTTGAATCAATCAATATTTTNA 492  
|  
|  
|  
QY 498 AACATTAATGATAGTCTCTGACATGATGATGATTTCTTTCTCTGTCAC 557  
|  
|  
|  
Db 493 AACATTAATGATAGTCTCTGACATGATGATGATTTCTTTCTCTGTCAC 552  
|  
|  
|  
QY 558 TTACCATGATGCTTAAATTAATGATCTATCA 590  
|  
|  
|  
Db 553 TTACCATGATGCTTAAATTAATGATCTATCA 585  
|  
|  
|

RESULT 4  
US-10-079-754A-6  
Sequence 6, Application US/10079754A  
Publication No. US20020164625A1

## GENERAL INFORMATION:

APPLICANT: Glenn, Matthew  
APPLICANT: Grigor, Murray R.  
APPLICANT: Molenaar, Adrian J.  
APPLICANT: Davis, Stephen R.  
TITLE OF INVENTION: Compositions Isolated from Bovine  
FILE REFERENCE: 11000.1068  
CURRENT APPLICATION NUMBER: US/10/079,754A  
CURRENT FILING DATE: 2002-02-19  
PRIOR APPLICATION NUMBER: US 09/699,146  
PRIOR FILING DATE: 2000-10-27  
PRIOR APPLICATION NUMBER: US 60,162,701  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: US 09/644,190  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: US 60,150,330  
PRIOR FILING DATE: 1999-08-23  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 525  
TYPE: DNA  
ORGANISM: Bovine  
US-10-079-754A-6

Query Match 70.1%; Score 423.4; DB 14; Length 525;  
Best Local Similarity 93.4%; Pred. No. 1.7e-85;  
Matches 466; Conservative 0; Mismatches 1; Indels 32; Gaps 1;

QY 138 CTGAGCTCCACCAATATGAAGATCTTTATCTTTGCTTCATTAAGGCTCTCATCTAGC 197  
|  
|  
|  
Db 27 CAGAGCTCCACCAATATGAAGATCTTTATCTTTGCTTCATTAAGGCTCTCATCTAGC 86  
|  
|  
|  
QY 198 CATGATTAGAGCTGATTCATCTGAAGAGAAAGCTCAAGGAAACGGAACAAATCATCAT 255  
|  
|  
|  
Db 87 CATGATTAGAGCTGATTCATCTGAAGAGAAAGCTCAAGGAAACGGAACAAATCATCAT 146  
|  
|  
|  
QY 256 -----AGAGATATTTTCAACAAATACAGCATAT 285  
|  
|  
|

Db 147 TGATAGTCTCCAGATTCCTTACTAATACAGAGATATTTTCAACAAATACAGCATAT 206  
|  
|  
|  
QY 286 CAACGATATCCATTAATTAATTCCTGCTGATCCATTCCTTAAATGCTGATAGTA 345  
|  
|  
|  
Db 207 CAACGATATCCATTAATTAATTCCTGCTGATCCATTCCTTAAATGCTGATAGTA 266  
|  
|  
|  
QY 346 CTACAGACATGATTAGAGATTTTTCACAAATGATTTTCTACTCTTTCTGTGTGT 405  
|  
|  
|  
Db 267 CTACAGACATGATTAGAGATTTTTCACAAATGATTTTCTACTCTTTCTGTGTGT 326  
|  
|  
|  
QY 406 GAAACCATCTTTCAATGAATTAACAAAGAAAAAATCAGTCAAGTGTGCA 465  
|  
|  
|  
Db 327 GAAACCATCTTTCAATGAATTAACAAAGAAAAAATCAGTCAAGTGTGCA 386  
|  
|  
|  
QY 466 CACATCTGGAATCAATATCATATTTTAAACATATATGATAGTCTGCACTAT 525  
|  
|  
|  
Db 387 CACATCTGGAATCAATATCATATTTTAAACATATATGATAGTCTGCACTAT 446  
|  
|  
|  
QY 526 GTAATGCTTCTACTTTCTTTCTGCTGCACTTACAGATGATGATTAATGATCT 585  
|  
|  
|  
Db 447 GTAATGCTTCTACTTTCTTTCTGCTGCACTTACAGATGATGATTAATGATCT 506  
|  
|  
|  
QY 586 ATCAAGCATAAAAA 604  
|  
|  
|  
Db 507 ATCAAGCATAAAAA 525  
|  
|  
|

RESULT 5  
US-10-079-754A-15  
Sequence 15, Application US/10079754A  
Publication No. US20020164625A1

## GENERAL INFORMATION:

APPLICANT: Glenn, Matthew  
APPLICANT: Grigor, Murray R.  
APPLICANT: Molenaar, Adrian J.  
APPLICANT: Davis, Stephen R.  
TITLE OF INVENTION: Compositions Isolated from Bovine  
FILE REFERENCE: 11000.1068  
CURRENT APPLICATION NUMBER: US/10/079,754A  
CURRENT FILING DATE: 2002-02-19  
PRIOR APPLICATION NUMBER: US 09/699,146  
PRIOR FILING DATE: 2000-10-27  
PRIOR APPLICATION NUMBER: US 60,162,701  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: US 09/644,190  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: US 60,150,330  
PRIOR FILING DATE: 1999-08-23  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15  
LENGTH: 267  
TYPE: DNA  
ORGANISM: Bovine  
US-10-079-754A-15

Query Match 39.5%; Score 238.4; DB 14; Length 267;  
Best Local Similarity 99.6%; Pred. No. 5e-44;  
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 138 CTGAGCTCCACCAATATGAAGATCTTTATCTTTGCTTCATTAAGGCTCTCATCTAGC 197  
|  
|  
|  
Db 28 CAGAGCTCCACCAATATGAAGATCTTTATCTTTGCTTCATTAAGGCTCTCATCTAGC 87  
|  
|  
|  
QY 198 CATGATTAGAGCTGATTCATCTGAAGAGAAAGCTCAAGGAAACGGAACAAATCATCAT 257  
|  
|  
|  
Db 88 CATGATTAGAGCTGATTCATCTGAAGAGAAAGCTCAAGGAAACGGAACAAATCATCAT 147  
|  
|  
|  
QY 258 AGATATTTTCAACAAATACAGCATATCAAGATATTCCTCTCTGCTGTA 317  
|  
|  
|  
Db 148 AGATATTTTCAACAAATACAGCATATCAAGATATTCCTCTCTGCTGTA 207  
|  
|  
|

Qy 318 TCCATTTCCTTAAAAATGCTGTAGTAACATCAGGACATGATTAGAGAGATTTTCAAA 377  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 208 TCATTTCCTTAAAAATGCTGTAGTAACATCAGGACATGATTAGAGAGATTTTCAAA 267

```

1 RESULT 6
2 US-10-079-623-200
3 ; Sequence 200, Application US/10079623
4 ; Publication No. US20020169302A1
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Havukkala, Ilkka J.
7 ; APPLICANT: Glenn, Matthew
8 ; APPLICANT: Grigor, Murray R.
9 ; APPLICANT: Molenaar, Adrian J.
10 ; TITLE OF INVENTION: Compositions isolated from bovine
11 ; TITLE OF INVENTION: Compositions isolated from bovine
12 ; FILE REFERENCE: 11000.10443
13 ; CURRENT APPLICATION NUMBER: US/10/079,623
14 ; CURRENT FILING DATE: 2002-02-19
15 ; SOFTWARE: FastSeq for Windows Version 4.0
16 ; SEQ ID NO 200
17 ; LENGTH: 267
18 ; TYPE: DNA
19 ; ORGANISM: Bovine
20 ; US-10-079-623-200

```

```

RESULT 7
US-10-079-754A-3
; Sequence 3, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions Isolated from Bovine
; TITLE OF INVENTION: Mammary Gland and Methods for Their Use
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079,754A
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/699,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60,162,701
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/644,190
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 60,150,330
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3

```

```

; LENGTH: 869
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-754A-3

```

Query Match	19.0%;	Score 115;	DB 14;	Length 869;
Best Local Similarity	92.4%;	Pred. No. 4.8e-16;		
Matches 121;	Conservative 0;	Mismatches 10;	Indels 0;	Gaps 0
QY	138	CTGCAGCTCCACCAATAATGAAGATCTTTATCTTTGTCTTCATTATAGCTTCATCCTAGC	197	
Db	56	CAGAGCTCCACCAATAATGAAGATCTTTATCTTTATCTTCATTATAGCTTCATCCTAGC	115	
QY	198	CATGATTTAGAGCTGATTCATCTGAAAGACAAACGTCAACGAAACCGAAAAAACATCATAG	257	
Db	116	CATGATTTAGAGCTGATTCATCTGAAAGACAAACGTCAACGAAACCGAAAAAACATCATAGT	175	
QY	258	ACGATATTTTC	268	
Db	176	ATGTATTCCTC	186	

```

RESULT 8
US-10-079-754A-5
, Sequence 5, Application US/10079754A
, Publication No. US20020164625A1
, GENERAL INFORMATION:
, APPLICANT: Glenn, Matthew
, APPLICANT: Grigor, Murray R.
, APPLICANT: Molenaar, Adrian J.
, APPLICANT: Davis, Stephen R.
, TITLE OF INVENTION: Compositions Isolated from Bovine
, TITLE OF INVENTION: Mammary Gland and Methods for Their Use
, FILE REFERENCE: 11000.1068
, CURRENT APPLICATION NUMBER: US/10/079,754A
, CURRENT FILING DATE: 2002-02-19
, PRIOR APPLICATION NUMBER: US 09/699,146
, PRIOR FILING DATE: 2000-10-27
, PRIOR APPLICATION NUMBER: US 60,162,701
, PRIOR FILING DATE: 1999-10-29
, PRIOR APPLICATION NUMBER: US 09/644,190
, PRIOR FILING DATE: 2000-08-22
, PRIOR APPLICATION NUMBER: US 60,150,330
, PRIOR FILING DATE: 1999-08-23
, NUMBER OF SEQ ID NOS: 15
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 5
, LENGTH: 96
, TYPE: DNA
, ORGANISM: Bovine
US-10-079-754A-5

```

RESULT 9  
US-09-992-600A-5  
; Sequence 5, Application US/09992600A  
; Publication No. US20030027161A1  
; GENERAL INFORMATION:  
; APPLICANT: Benjamin, Stephanie  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

```
FILE REFERENCE: 91.US4.DIV
CURRENT APPLICATION NUMBER: US/09/992,600A
CURRENT FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 114
SOFTWARE: JPatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
FEATURE:
NAME/KEY: CDS
LOCATION: 84..317
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 318..438
FEATURE:
NAME/KEY: polyA signal
LOCATION: 397..402
FEATURE:
NAME/KEY: polyA site
LOCATION: 423..438
US-09-992-600A-5
```

```
Query Match 11.7% Score 70.4; DB 11; Length 438;
Best Local Similarity 69.9%; Pred. No. 3.8e-06;
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY 120 GAATTCATCTTTCATGACTGACCTCCACCAATATGAAGATCTTATCTTGTCTTCAT 179
DB 50 GACTCTCCTCTTGAGTAAGAGACTCAGCACTATGAAGTTTGTCTTGTCTTGTAGT 109
QY 180 TATGCTCTCATCTAGCCATGATTAGAGCTGATTCATCTGAAGAAAGTCACAGAA 239
DB 110 CTGGCTCTCATGATTTCCATGATTAGCGCTGATTCACATGAAGAGACATCATGGGTA 169
QY 240 ACGGAAAAACATCAT 255
DB 170 TAGAAGAAAAATTCAT 185

RESULT 10
US-09-924-340-5
Sequence 5, Application US/09924340
Publication No. US20030027248A1
GENERAL INFORMATION:
APPLICANT: Benjamin, Stephanie
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US2.REG
CURRENT APPLICATION NUMBER: US/09/924,340
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
```

```
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: JPatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
NAME/KEY: CDS
LOCATION: 84..317
NAME/KEY: 3'UTR
LOCATION: 318..438
NAME/KEY: polyA signal
LOCATION: 397..402
NAME/KEY: polyA site
LOCATION: 423..438
US-09-924-340-5
```

```
Query Match 11.7% Score 70.4; DB 11; Length 438;
Best Local Similarity 69.9%; Pred. No. 3.8e-06;
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY 120 GAATTCATCTTTCATGACTGACCTCCACCAATATGAAGATCTTATCTTGTCTTCAT 179
DB 50 GACTCTCCTCTTGAGTAAGAGACTCAGCACTATGAAGTTTGTCTTGTCTTGTAGT 109
QY 180 TATGCTCTCATCTAGCCATGATTAGAGCTGATTCATCTGAAGAAAGTCACAGAA 239
DB 110 CTGGCTCTCATGATTTCCATGATTAGCGCTGATTCACATGAAGAGACATCATGGGTA 169
QY 240 ACGGAAAAACATCAT 255
DB 170 TAGAAGAAAAATTCAT 185
```

```
RESULT 11
US-09-992-095B-5
Sequence 5, Application US/0992095B
Publication No. US20030157485A1
GENERAL INFORMATION:
APPLICANT: Tanaka, Hiroaki
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US5.DIV
CURRENT APPLICATION NUMBER: US/09/992,095B
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: JPatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
FEATURE:
NAME/KEY: CDS
LOCATION: 84..317
FEATURE:
```

NAME/KEY: 3'UTR  
LOCATION: 318..438  
FEATURE:  
NAME/KEY: polyA\_signal  
LOCATION: 397..402  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 423..438  
US-09-992-095B-5

Query Match 11.7%; Score 70.4; DB 13; Length 438;  
Best Local Similarity 69.9%; Pred. No. 3.8e-06;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACGACCTCCACCAATATGAGATCTTTATCTTTGCTTCAT 179  
DB 50 GACTCTCTCTTGTAGTAAAGACTCAGCCACATATGAAGTTTGTCTTTGCTTAACT 109  
QY 180 TATGGCTCTCATCTTACCATGATTAGAGCTGATTCATCTGAAGAGAAAGTCACAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTAGCGGTGATTCACATGAAGAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAAGAAATTCAT 185

RESULT 12  
US-10-154-678-5  
Sequence 5, Application US/10154678  
Publication No. US20030162186A1  
GENERAL INFORMATION:  
APPLICANT: Benjamin, Stephane  
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
FILE REFERENCE: 182.USI.REG  
CURRENT FILING DATE: 2002-10-15  
PRIOR FILING DATE: 2002-10-15  
PRIOR APPLICATION NUMBER: US 09/924,340  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/305,456  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/302,277  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/298,698  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/293,574  
PRIOR FILING DATE: 2001-05-25  
NUMBER OF SEQ ID NOS: 112  
SOFTWARE: JPatent  
SEQ ID NO 5  
LENGTH: 438  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..83  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 84..317  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 318..438  
FEATURE:  
NAME/KEY: polyA\_signal  
LOCATION: 397..402  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 423..438  
US-10-154-678-5

Query Match 11.7%; Score 70.4; DB 13; Length 438;  
Best Local Similarity 69.9%; Pred. No. 3.8e-06;

Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
QY 120 GAATTCATCTTTCATGACGACCTCCACCAATATGAGATCTTTATCTTTGCTTCAT 179  
DB 50 GACTCTCTCTTGTAGTAAAGACTCAGCCACATATGAAGTTTGTCTTTGCTTAACT 109  
QY 180 TATGGCTCTCATCTTACCATGATTAGAGCTGATTCATCTGAAGAGAAAGTCACAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTAGCGGTGATTCACATGAAGAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAAGAAATTCAT 185

RESULT 13  
US-09-999-570-5  
Sequence 5, Application US/09999570  
Publication No. US20030170628A1  
GENERAL INFORMATION:  
APPLICANT: Benjamin, Stephane  
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
FILE REFERENCE: G-091US08DIY  
CURRENT FILING DATE: 2001-06-14  
PRIOR FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 09/924,340  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: PCT/IB01/01715  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/305,456  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/302,277  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/298,698  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/293,574  
PRIOR FILING DATE: 2001-05-25  
NUMBER OF SEQ ID NOS: 112  
SOFTWARE: JPatent  
SEQ ID NO 5  
LENGTH: 438  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..83  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 84..317  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 318..438  
FEATURE:  
NAME/KEY: polyA\_signal  
LOCATION: 397..402  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 423..438  
US-09-999-570-5

Query Match 11.7%; Score 70.4; DB 13; Length 438;  
Best Local Similarity 69.9%; Pred. No. 3.8e-06;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACGACCTCCACCAATATGAGATCTTTATCTTTGCTTCAT 179  
DB 50 GACTCTCTCTTGTAGTAAAGACTCAGCCACATATGAAGTTTGTCTTTGCTTAACT 109  
QY 180 TATGGCTCTCATCTTACCATGATTAGAGCTGATTCATCTGAAGAGAAAGTCACAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTAGCGGTGATTCACATGAAGAGACATCATGGGTA 169

QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAAGAAAATTCAT 185

## RESULT 14

US-10-001-142-5  
Sequence 5, Application US/10001142  
Publication No. US20030198954A1  
GENERAL INFORMATION:

APPLICANT: Benjamin, Stephanie  
APPLICANT: Tanaka, Hiroaki  
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
FILE REFERENCE: 91.US7.DIV  
CURRENT APPLICATION NUMBER: US/10/001,142  
CURRENT FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: US 09/924,340  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: PCT/IB01/01715  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/305,456  
PRIOR FILING DATE: 2001-07-13  
NUMBER OF SEQ ID NOS: 112  
SOFTWARE: JPatent  
SEQ ID NO 5  
LENGTH: 438  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..83  
NAME/KEY: CDS  
LOCATION: 84..317  
NAME/KEY: 3'UTR  
LOCATION: 318..438  
NAME/KEY: polyA\_signal  
LOCATION: 397..402  
NAME/KEY: polyA\_site  
LOCATION: 423..438  
US-10-001-142-5

## Query Match

Best Local Similarity 11.7%; Score 70.4; DB 13; Length 438;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTCATGACTGAGCTCCACCAATATGAAGATCTTATCTTGTCTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGAGCTCAGCCACTATGAAGTTTGTCTTGTCTTGTAGT 109  
QY 180 TATGCTCTCATCTAGCCATGATTTAGAGCTGATTCATCTGAAGAAAGTCACAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTTAGCGCTGATTCATGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAAGAAAATTCAT 185

## RESULT 15

US-10-000-489-5  
Sequence 5, Application US/10000489  
Publication No. US20030092011A1  
GENERAL INFORMATION:

APPLICANT: Benjamin, Stephanie  
APPLICANT: Tanaka, Hiroaki  
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
FILE REFERENCE: 91.US6.DIV  
CURRENT APPLICATION NUMBER: US/10/000,489  
CURRENT FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: US 09/924,340  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: PCT/IB01/01715  
PRIOR FILING DATE: 2001-08-06

PRIOR APPLICATION NUMBER: US 60/305,456  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/302,277  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/298,698  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/293,574  
PRIOR FILING DATE: 2001-05-25  
NUMBER OF SEQ ID NOS: 112  
SOFTWARE: JPatent  
SEQ ID NO 5  
LENGTH: 438  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..83  
NAME/KEY: CDS  
LOCATION: 84..317  
NAME/KEY: 3'UTR  
LOCATION: 318..438  
NAME/KEY: polyA\_signal  
LOCATION: 397..402  
NAME/KEY: polyA\_site  
LOCATION: 423..438  
US-10-000-489-5

## Query Match

Best Local Similarity 11.7%; Score 70.4; DB 15; Length 438;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTCATGACTGAGCTCCACCAATATGAAGATCTTATCTTGTCTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGAGCTCAGCCACTATGAAGTTTGTCTTGTCTTGTAGT 109  
QY 180 TATGCTCTCATCTAGCCATGATTTAGAGCTGATTCATCTGAAGAAAGTCACAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTTAGCGCTGATTCATGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAAGAAAATTCAT 185

Search completed: December 13, 2003, 03:06:26  
Job time : 284 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2003, 01:35:43 ; Search time 1865 Seconds  
(without alignments)  
7871.262 Million cell updates/sec

Title: US-10-079-754a-4

Perfect score: 604  
Sequence: 1 gaagatatttcagttctata.....tatcaagcataaaaaaaaaa 604

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estbua:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_ges\_hum:\*  
18: em\_ges\_inv:\*  
19: em\_ges\_pln:\*  
20: em\_ges\_vrt:\*  
21: em\_ges\_fun:\*  
22: em\_ges\_mam:\*  
23: em\_ges\_mus:\*  
24: em\_ges\_pro:\*  
25: em\_ges\_rod:\*  
26: em\_ges\_pbg:\*  
27: em\_ges\_vrl:\*  
28: gb\_ges1:\*  
29: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query length	DB ID	Description
1	82.2	13.6	553	14	CB985810
2	78.6	13.0	546	14	CB957544
3	78.6	13.0	555	14	CB986888
4	78.6	13.0	792	14	CB958289

5	78.6	13.0	793	14	CB984475	CB984475
6	78.6	13.0	831	14	CB985521	CB985521
7	78.6	13.0	929	14	CB987745	CB987745
8	77.6	12.8	583	2	HSN076012	Bx485825 Homo sapi
9	77.6	12.8	622	2	HSN075901	Bx485714 Homo sapi
10	77.6	12.8	773	14	CB987175	CB987175
11	77	12.7	312	9	AA376700	AA376700 EST89157
12	77	12.7	327	9	AA382524	AA382524 EST95743
13	77	12.7	330	9	AA376704	AA376704 EST89161
14	77	12.7	348	9	AA376689	AA376689 EST89145
15	77	12.7	390	13	BX283872	BX283872
16	77	12.7	527	14	CB986478	CB986478
17	77	12.7	533	14	CB985159	CB985159
18	77	12.7	540	14	CB957853	CB957853
19	77	12.7	543	14	CB958523	CB958523
20	77	12.7	543	14	CB987834	CB987834
21	77	12.7	544	14	CB957223	CB957223
22	77	12.7	544	14	CB959166	CB959166
23	77	12.7	545	14	CB958545	CB958545
24	77	12.7	545	14	CB959029	CB959029
25	77	12.7	546	14	CB956307	CB956307
26	77	12.7	546	14	CB956449	CB956449
27	77	12.7	546	14	CB957121	CB957121
28	77	12.7	546	14	CB957972	CB957972
29	77	12.7	546	14	CB985184	CB985184
30	77	12.7	546	14	CB986281	CB986281
31	77	12.7	546	14	CB986417	CB986417
32	77	12.7	546	14	CB987123	CB987123
33	77	12.7	546	14	CB987595	CB987595
34	77	12.7	547	14	CB955720	CB955720
35	77	12.7	547	14	CB956188	CB956188
36	77	12.7	547	14	CB956487	CB956487
37	77	12.7	547	14	CB956536	CB956536
38	77	12.7	547	14	CB956860	CB956860
39	77	12.7	547	14	CB956877	CB956877
40	77	12.7	547	14	CB957224	CB957224
41	77	12.7	547	14	CB957280	CB957280
42	77	12.7	547	14	CB957414	CB957414
43	77	12.7	547	14	CB957440	CB957440
44	77	12.7	547	14	CB957523	CB957523
45	77	12.7	547	14	CB957940	CB957940

## ALIGNMENTS

RESULT 1  
LOCUS CB985810  
DEFINITION AGENCOURT 13647469 NIH MGC 184 Homo sapiens cDNA clone  
IMAGE:30329983 5', mRNA sequence.  
ACCESSION CB985810  
VERSION CB985810.1 GI:30280334  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 553)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strussberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.lnl.gov>  
Plate: NDCM143 row: b column: 08  
High quality sequence stop: 553.

**FEATURES**  
**source**

FEATURES	SOURCE	Location/Qualifiers
1..553		location="Homo sapiens"
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone_image="30329983"
		/lab_host="DH10B (T1 phage-resistant)"
		/clone_id="NIH_MGC_184"
		/note="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site_1: Sfil (ggccattatggcc); Site_2: Sfil (ggccggccggcc)"
		Library is oligo-dT primed and directionally cloned. cDNA was prepared from a glandular pool of tissues from thyroid parathyroid, adrenal, cortex and pineal gland. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCACATTAGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGAGCCGACGCGCCACCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.38 kb (range 0.60-3.5 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
171 a	117 c	81 g 183 t 1 others

**Source**

```

source
1. .546
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30352504"
/lab_host="DH10B (TI phage-resistant)"
/clone_1b="NIH MGC 184"
/notes="Organ: Pooled-Glandular; Vector: pMDR-LIB; Site_1: Sfil (ggccatratggcc); Site_2: Sfil (ggccgctcgcc); cdna library is oligo-dt primed and directionally cloned. cdna was prepared from a glandular pool of tissues from thyroid, parathyroid, adrenal, cortex and pineal gland. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGAGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCCGCAGCGCCGCACATG-dt(30)-BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.38 kb (range 0.60-3.5 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."
BASE COUNT
166 a 103 c 87 g 189 t 1 others
ORIGIN
Query Match 13.0%; Score 78.6; DB 14; Length 546;
Best Local Similarity 60.2%; Pred. No. 0.086;
Matches 151; Conservative 0; Mismatches 94; Indels 6; Gaps 1;
QY 120 GAATTCATCTTTCATGACTGAGCTCCACCAATATGAGATCTTTATCTTTCAT 179
DB 41 GACTCTCCTTGAGTAAAGAGACTCAGCAACATGATGATCTTGTCTTCTTAGT 100
QY 180 TATGGCTTCATCTTACCATGATTAAAGCTGATTCATCGAAGAGAACGACAGAA 239
DB 101 CTGGCTCTCATGATTTCCATGATTTACCGCTGATTCACATGAAAAGACATCTGGATA 160
QY 240 ACGAAAAAATCATATAGAGATTTTTCACAACATACAGCCATATACAGATATCCAT 299
DB 161 TAGAAGAAATTCAT-----GAAAGCATCATTCACATCGAATTTTCATTTATGG 214
QY 300 AAATTATCTCTCGTATCCATTTCTTAAATGCTGCTTAGTAACTACAGACATGAT 359
DB 215 GGACTATGATCAATATATCATATATGACAATGATATCTTAGTATATCATCGGGCATGAT 274
QY 360 TAGAGAGATT 370
DB 275 TATAGAGATT 285
RESULT 3
LOCUS CB986888
DEFINITION AGNCOURT 13650592 NIH MGC 184 Homo sapiens CDNA clone
IMAGE:30329446 5', mRNA sequence.
ACCESSION CB986888
VERSION CB986888.1 GI:30281408
KEYWORDS EST.
ORGANISM Homo sapiens (human)
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 555)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov

```



Plate: NDCM141 row: k column: 23  
High quality sequence stop: 551.

**FEATURES**

**SOURCE**

```

1. .555
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30329446"
/lab_host="MDH10B (T1 phage-resistant)"
/clone_lib="NIH MCCLIB 184"
/note="Organ: Pooled Glandular; Vector: pDNR-LIB; Site: 1:
      S11 (ggcagcttggcc); Site: 2: S11 (ggccgccttgcgc);
      Library is oligo-dT primed and directionally cloned. cDNA
      was prepared from a glandular pool of tissues from thyroid
      parathyroid, adrenal, cortex and pineal gland. 5' and 3'
      adaptors were used in cloning as follows: 5' adaptor
      sequence: 5'-CACGCGATGATGCGC-3' and 3' adaptor sequence:
      5'-ATTCTAGAGCGCGCGCGCGCATGATG-dT(30)BN-3' (where B = A,
      C, or G and N = A, C, or T). Average insert size 1.38
      kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
      by PCR. This library was enriched for full-length clones
      and was constructed by Clontech Laboratories (Palo Alto,
      CA). Note: This is a NIH MCCLIB Library."

```

BASE COUNT	169 a	104 c	88 g	190 t	4 others
ORIGIN					

Query Match	13.0%;	Score 78.6;	DB 14;	Length 555;
Best Local Similarity	60.2%;	Pred. No. 0.085;		
Matches 151;	Conservative 0;	Mismatches 94;	Indels 6;	Gaps 1

Qy	120	GAATTCATCTTTCATGACTGACCTCACCAAAATGAAAGATCTTATCTTGGCTTCAT	179
Db	41	GACTCTCCTCTTGAGTAAAGACCTCAGCAACTGAAAGTTTGTGGCTTTGCTTTAGT	100
Qy	180	TATGCGCTCTCATCTTAGCCATGATTTAGAGCTGATTCATCTGAGAGAAAGCTCACAGAA	239
Db	101	CTTGGGCTCTCATTAATTTCCATGATATTAGCGCTCATTCACATAAAAGAGACATCTAGGGTA	160
Qy	240	ACGGAATAAAACATCATAGAGAGATATTTTGAACATACAGCGCATATCAAGATATCCACT	299
Db	161	TAGAAAGAAAATTCAT-----GAAAAGCATATTCACATGAGAAATTTCCATTTTATGG	214
Qy	300	AAATTAATCCTCGCGATTCATTTCTTTAAATGCTGTTAGTACTACAGACATGAT	359
Db	215	GGAGCTATGAGACCAATTTATCTATATGACAAATTGATATCTTAGTATCATGGGCGCATGAT	274
Qy	360	TAGAGAGATTT	370
Db	275	TATAGAGTTT	285

RESULT 4				
CB958289				
LOCUS	CB958289	792 bp	mRNA	linear EST 29-APR-2001
DEFINITION	AGNCNC8289			
	AGNCNC8289	13778124	NIH_MGC_184 Homo sapiens	clone
	IMAGE:30352112 5',			mRNA sequence.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 792)	NIH-MGC	<a href="http://mgc.nhl.nih.gov/">http://mgc.nhl.nih.gov/</a>	National Institutes of Health, Mammalian Gene Collection	(MGC)
	Unpublished	Contact: Robert Strausberg, Ph.D.		

clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
plate: NDCM50 row: 1 column: 09  
High quality sequence step: 498.  
Location/Qualifiers  
1..792

BASE COUNT	247 a	165 c	144 g	235 t	1 other
ORIGIN					

Query Match	13.0%;	Score 78.6;	DB 14;	Length 792;
Best Local Similarity	60.2%;	Pred. No. 0.074;		
Matches 151; Conservative	0;	Mismatches . 94;	Indels 6;	Gaps 1

Qy	120	GAATTCATCTTCACTGAGTCCACCAAAATGAAATCTTATCTTGGCTTCAT	179
Db	40	GACTCCCTCTTGAGTAAAGGACCTCACCACACATGAAATTTTCTCTTGGTTAGT	99
Qy	180	TATGGCTTCATCTTAGCCATGATTTAGAGCTGATTCATCTGAGAGAAAGTCAAGAA	239
Db	100	CTTGGCTTCATGATTTCCATGATTAAGGCTGATTCACATGAAAAGAGATATGGGTA	159
Qy	240	ACGGAAGAAACATCATAGAGATATTTTCAACATPACAGCATATCAAGATATCCACT	299
Db	160	TAGAGAAAATTCATGAAAAA-----CATCATTCACATCGAATTCATTTTATGG	213
Qy	300	AAATTATCCCTCGGATTCATTTCTTAAATGCTGTTAGTACACAGACATGAT	359
Db	214	GGACTATGATCAAAATTATCTATATGACAAATTGATATTCCTTAGTAATCATGGGGCATGAT	273
Qy	360	TAGAGATTT	370
Db	274	TATAGAGTTT	284

RESULT 5	
CB984475	
LOCUS	CB984475
DEFINITION	793 bp mRNA linear EST 01-MAY-2001
IMAGE:30327061.5'	AGNCCOURT 13459298 NIH MGC 184 Homo sapiens cDNA clone mRNA sequence.

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs-rc@mail.nih.gov Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits



**AUTHORS** NIH-MGC <http://mgc.nci.nih.gov/>.  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished  
**COMMENT** Contact: Robert Strausberg, Ph.D.

BASE COUNT	269 a	199 c	185 g	274 t	2 others
ORIGIN					

Query Match	13.0%;	Score 78.6;	DB 14;	Length 929;
Best Local Similarity	60.2%;	Pred. No. 0.069;		
Matches 151; Conservative	0;	Mismatches 94;	Indels 6;	Gaps 1

RESULT 8	
HSM076012	
ID	HSM076012
	standard; RNA; EST; 583 BP

XX	
DE	Homo sapiens mRNA; EST DKFZp686I09248_r1 (from clone DKFZp686I09248)
XX	
KW	EST; expressed sequence tag.
XX	

XX	
DT	09-MAY-2003 (Rel. 75, Created)
DT	09-MAY-2003 (Rel. 75, Last updated, Version 1)

RESULT 9	
MSM075901	
ID	MSM075901
	standard; RNA; EST; 622 BP.

```

XX AC BX485714;
XX SV BX485714.1
XX DT 09-MAY-2003 (Rel. 75, Created)
XX DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX DE Homo sapiens mRNA; EST DKFZp686E20248_r1 (from clone DKFZp686E20248)
XX KM EST; expressed sequence tag.
XX OS Homo sapiens (human)
XX OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
XX OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
XX [1]
XX RP 1-622
XX RA Othenwajder B., Obermaier B., Deutschenbaur S., Mewes H.W., Weil B.,
XX RA Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
XX RT Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
XX RL MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY.
XX CC This is the 5' sequence of the clone insert
XX CC Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
XX CC Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
XX CC sequenced by MediGenomix (Martinsried/Germany) within the cDNA
XX CC sequencing consortium of the German Genome Project.
XX CC No st sequence available.
XX CC This clone (DKFZp686E20248) is available at the RZPD in Berlin.
XX CC Please contact the RZPD: Ressourcenzentrum, Heubweg 6,
XX CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX FH Key Location/Qualifiers
XX FT source 1..622
XX FT /db_xref="taxon:9606"
XX FT /mol_type="mRNA"
XX FT /organism="Homo sapiens"
XX FT /clone="DKFZp686E20248"
XX FT /clone_id="686 (synonym: h1cc3). Vector pSport1_sfi; host
XX FT DH10B; sites SfiI + SfiIB"
XX FT /dev_stage="adult"
XX FT /tissue_type="CDNA-collection"
XX SQ Sequence 622 BP; 161 A; 103 C; 128 G; 229 T; 1 other;
Query Match 12.8%; Score 77.6; DB 2; Length 622;
Best Local Similarity 56.2%; Pred. No. 0.11;
Matches 189; Conservative 0; Mismatches 139; Indels 8; Gaps 2;
QY 120 GAATTCATCTTATGACGAGTCCACCAATATGAAATCTTTATCTTCTTCAT 179
DB 36 GACCTCTCTTGAGTAAAGAGCTCAAGCCACTATGAAATTTTGTCTTGTAGT 95
QY 180 TATGGCTCTCATCTAGCATGATTAGAGCTGATTCACTGGAAGAAAGCTCAGCAA 239
DB 96 CTGGCTCTCTATGATTTCCATGATTAAGGCTGATTCATGAAAGAGACATCGGTA 155
QY 240 ACGGAAAAAATCATGAGAGATTTTCAACAATATACAGCATATCAAGATATCCACT 299
DB 156 TAGAAGAAATTCAT-----GAAAAGCATTCATCAGATGAGAAATTTTCATTATGG 209
QY 300 AAATTATCTCTCGGATGATTCCTTAAATGCTGCTTAGTAATCAGAGCATGAT 359
DB 210 GGAATATGATCAAAATATCATATGACAATGATATCTTAGTAATCATCGGCGCATGAT 269
QY 360 TAGAGATTTTTCACATGATTTTTCCTACTCTTCTGTTGTGTGTAAGAAACATCTTTC 419
DB 270 TATAGAG--GTAGCTGATCTAGTACTGTCTTCTTGAAGATGTCAACCTGACAGTT 327
QY 420 AAATGATTAATAACAAAGAAAAAATATCATGTAAGT 455

```

```

DB 328 TAAAGAAAAAGCATAGCTTAACAAACCATTCACGT 363
RESULT 10
CB987175 773 bp mRNA linear EST 01-MAY-2003
LOCUS
DEFINITION
AGENCOURT 13650773 NIH MGC 184 Homo sapiens cDNA clone
IMAGE:30329891 5', mRNA sequence.
CB987175
ACCESSION
CB987175.1 GI:30281695
VERSION
CB987175.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 773)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
JOURNAL
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palovics
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: NDCM142 row: n column: 12
High quality sequence stop: 181.
FEATURES
source
1..773
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30329891"
/lab_host="DH10B (T1 phage-resistant)"
/clone_id="NIH MGC 184"
/note="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site_1:
SfiI (ggccatcggcc); Site_2: SfiI (ggccgctggcc);
Library is oligo-dT primed and directionally cloned. cDNA
was prepared from a glandular pool of tissues from thyroid,
parathyroid, adrenal, cortex and pineal gland. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCGCATATGACC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGAGGCGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.38
kb (range 0.60-3.5 kb). 15/15 clones contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC library."
BASE COUNT 227 a 163 c 159 g 223 t 1 others
ORIGIN
Query Match 12.8%; Score 77.6; DB 14; Length 773;
Best Local Similarity 59.8%; Pred. No. 0.11;
Matches 150; Conservative 0; Mismatches 95; Indels 6; Gaps 1;
QY 120 GAATTCATCTTATGACGAGTCCACCAATATGAAATCTTTATCTTCTTCAT 179
DB 41 GACCTCTCTTGAGTAAAGAGCTCAGCCACTATGAAATTTTGTCTTGTAGT 100
QY 180 TATGGCTCTCATCTAGCATGATTAGAGCTGATTCACTGGAAGAAAGCTCAGCAA 239
DB 101 CTGGCTCTCATATTTCCATGATTAAGGCTGATTCACAGAAAGAGACATATGGTA 160
QY 240 ACGGAAAAAATCATGAGAGATTTTCAACAATATACAGCATATCAAGATATCCACT 299
DB 161 TAGAAGAAATTCAT-----GAAAAGCATTCATCAGATGAGAAATTTTCATTATGG 214
QY 300 AAATTATCTCTCGGATGATTCCTTAAATGCTGCTTAGTAATCAGAGCATGAT 359

```





TITLE Venter, J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
JOURNAL Nature 377 (6547 Suppl.), 3-174 (1995)  
MEDLINE 96026280  
PubMed 7566098  
COMMENT Other\_ESTS: EST89144 THC102726  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/cdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
source  
Location/Qualifiers  
1..348  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="ATCC (inhost):181112"  
/db\_xref="taxon:9606"  
/dev\_stage="adult"  
/clone\_lib="Salivary gland"  
/note="Organ: salivary gland; Vector: pBluescript SK-;  
Site 1: EcoRI; Site 2: XhoI"  
BASE COUNT 103 a 66 c 60 g 119 t  
ORIGIN

Query Match 12.7%; Score 77; DB 9; Length 348;  
Best Local Similarity 59.8%; Pred. No. 0.17;  
Matches 150; Conservative 0; Mismatches 95; Indels 6; Gaps 1;

QY 120 GAATTCATCTTTCATGACTGACTCCACCAATATGAAATCTTATCTTGTCTTCA 179  
DB 12 GACTCTCCTCTGATTAAGAGCTCAGCAGCTATGAAATTTTGTCTTGTCTTCA 71  
QY 180 TATGCTCTCATCCCTAGCATATATGAGCTATCTATCTGTAAGAAAGCTCAGAGAA 239  
DB 72 CTGGCTCTCATGATTTCCATGATTAAGGCTGATTCACATGAAAGAGCATCAGGGTA 131  
QY 240 ACGGAAAAACATCATAGAGATATTTTCAACATATCAGCATATCAAGATATCCACT 299  
DB 132 TAGAAGAAATTCAT-----GAAAGCATCATTCACATCGAATTTCCATTTATGG 185  
QY 300 AAATTATCCTCTGCGTATTCATTTCTTAAATGCTGTTAGTAATACTACAGACATGAT 359  
DB 186 GGACTATGATCAATTTATCTATATGACATTTGATATCTTATATCATGCGGATGAT 245  
QY 360 TAGAGAGATTT 370  
DB 246 TATAGAGGTTT 256

RESULT 15  
BX283872 390 bp mRNA linear EST 05-MAR-2003  
LOCUS BX283872 NIH MGC 81 Homo sapiens cDNA IMAGE:4289874, mRNA sequence.  
DEFINITION IMAGE:4289874, mRNA sequence.  
ACCESSION BX283872  
VERSION BX283872.1 GI:28848326  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 390)  
AUTHORS Ebert, U., Heli, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,  
Radelof, U., Schneider, D. and Korn, B.  
TITLE Human Unigeneset - RZPD3  
JOURNAL Unpublished

## COMMENT

Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
RZPD; IMGP958N191131.  
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;  
Human Unigeneset - RZPD3 (RZPDLIB No. 972)  
<http://www.rzpd.de/CloneCards/cgi-bin/showlib.pl.cgi/response?libNo=972> Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de  
This clone is available royalty-free from RZPD;  
contact RZPD ([clone@rzpd.de](mailto:clone@rzpd.de)) for further information. Seq primer:  
pcmv-m13v. Primer sequence: CATTATAACAGCGCCAGT.

FEATURES  
source  
Location/Qualifiers  
1..390  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMGP958N191131 ; IMAGE:4289874"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH MGC 81"  
/note="Organ: muscle (skeletal); Vector: pDNR-LIB  
(Clontech); Site 1: SfiI (ggcgccctcgccg); Site 2: SfiI  
(ggcattatggcc); 5' and 3' adaptors were used in cloning  
as follows: 5' adaptor sequence: 5'-CACGGCCATTTATGGCC-3'  
and 3' adaptor sequence:  
5'-ATCTAGAGCGGAGCGCGCCAGCATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size  
1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained  
inserts by PCR. This library was enriched for full-length  
clones and was constructed by Clontech Laboratories (Palo  
Alto, CA)."

BASE COUNT 120 a 75 c 66 g 129 t  
ORIGIN  
Query Match 12.7%; Score 77; DB 13; Length 390;  
Best Local Similarity 59.8%; Pred. No. 0.16;  
Matches 150; Conservative 0; Mismatches 95; Indels 6; Gaps 1;

QY 120 GAATTCATCTTTCATGACTGACTCCACCAATATGAAATCTTATCTTGTCTTCA 179  
DB 35 GACTCTCCTCTGATTAAGAGCTCAGCAGCTATGAAATTTTGTCTTGTCTTCA 94  
QY 180 TATGCTCTCATCCCTAGCATATATGAGCTATCTATCTGTAAGAAAGCTCAGAGAA 239  
DB 95 CTGGCTCTCATGATTTCCATGATTAAGGCTGATTCACATGAAAGAGCATCAGGGTA 154  
QY 240 ACGGAAAAACATCATAGAGATATTTTCAACATATCAGCATATCAAGATATCCACT 299  
DB 155 TAGAAGAAATTCAT-----GAAAGCATCATTCACATCGAATTTCCATTTATGG 208  
QY 300 AAATTATCCTCTGCGTATTCATTTCTTAAATGCTGTTAGTAATACTACAGACATGAT 359  
DB 209 GGACTATGATCAATTTATCTATATGACATTTGATATCTTATATCATGCGGATGAT 268  
QY 360 TAGAGAGATTT 370  
DB 269 TATAGAGGTTT 279

Search completed: December 13, 2003, 03:00:29  
Job time : 1867 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2003, 22:51:27 ; Search time 2599 Seconds  
(without alignments)  
9507.278 Million cell updates/sec

Title: US-10-079-754A-4  
Perfect score: 604  
Sequence: 1 gaagatatttcagttcctata.....catcaagcataaaaaaaaaa 604

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: GenEmbl:\*
- 2: gb\_ba:\*
- 3: gb\_hcg:\*
- 4: gb\_in:\*
- 5: gb\_ov:\*
- 6: gb\_ov:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_on:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_hcg\_hum:\*
- 31: em\_hcg\_inv:\*
- 32: em\_hcg\_other:\*
- 33: em\_hcg\_mus:\*
- 34: em\_hcg\_pln:\*
- 35: em\_hcg\_rtd:\*
- 36: em\_hcg\_mam:\*
- 37: em\_hcg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_hcgo\_hum:\*
- 40: em\_hcgo\_mus:\*
- 41: em\_hcgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	141	23.3	195534	2 AC134173	AC134173 Bos tauru
2	77	12.7	566	2 BC017835	BC017835 Homo sapi
3	75.4	12.5	357	6 BD071675	BD071675 Secreted
4	70.4	11.7	438	6 AX588147	AX588147 Sequence
5	70.4	11.7	438	6 AX588483	AX588483 Sequence
6	68.2	11.3	438	6 AX616321	AX616321 Sequence
7	68.2	11.3	524	6 HMH152X	M2665 Human histra
8	68.2	11.3	558	9 BC009791	BC009791 Homo sapi
9	67.6	11.2	542	9 HUMSTRNA	M18371 Human stach
10	67.6	11.2	552	9 HUMSTATHA	M18078 Human stach
11	67.6	11.2	1584	9 AK092678	AK092678 Homo sapi
12	67.2	11.1	203	6 BD071749	BD071749 Secreted
13	66.6	11.0	480	4 HMH151X	M2664 Human histra
14	65.6	10.9	486	4 AY154893	AY154893 Bos tauru
15	62.6	10.4	491	9 HUMBRPA	M18372 Human histri
16	60.8	10.1	516	6 BD071674	BD071674 Secreted
C 17	60.4	10.0	110000	2 AC106950_1	Continuation (2 of
18	60.4	10.0	231767	2 AC094531	AC094531 Rattus no
19	60.4	10.0	240446	2 AC131219	AC131219 Rattus no
20	59.2	9.8	7550	9 HMH15102	L04132 Human histra
21	59.2	9.8	141568	9 AC063956	AC063956 Homo sapi
22	59.2	9.8	185969	2 AC069037	AC069037 Homo sapi
23	56.6	9.4	143842	2 AC134934	AC134934 Bos tauru
24	56.6	9.4	195534	2 AC134173	AC134173 Bos tauru
25	55.8	9.2	176	6 BD071751	BD071751 Secreted
C 26	55.6	9.2	218675	2 AC015847	AC015847 Homo sapi
27	55	9.1	249943	3 AE014823	AE014823 Plasmodiu
C 28	54.6	9.0	49306	3 AC115606	AC115606 Dictyoste
29	54.4	9.0	253001	3 AE014834	AE014834 Plasmodiu
C 30	54.2	9.0	154071	3 AC115598	AC115598 Dictyoste
C 31	54.2	9.0	250823	3 AE014821	AE014821 Plasmodiu
C 32	53.6	8.9	13449	6 AX346287	AX346287 Sequence
C 33	53.6	8.9	250022	3 AE014824	AE014824 Plasmodiu
C 34	52.4	8.7	82139	3 AC115684	AC115684 Dictyoste
C 35	52	8.6	186834	5 BX004887	BX004887 Zebrachia
C 36	52	8.6	349980	6 AX344567	AX344567 Sequence
C 37	51.8	8.6	254436	3 AE014827	AE014827 Plasmodiu
C 38	51.6	8.5	890	3 EHAPCG	X76903 E. histolyti
C 39	51.6	8.5	7736	3 AF310889	AF310889 Dictyoste
C 40	51.6	8.5	36148	3 AC116100	AC116100 Dictyoste
C 41	51.6	8.5	110000	2 PFMALBP1_10	Continuation (11 o
C 42	51.4	8.5	161549	2 AC024676	AC024676 Homo sapi
C 43	51.2	8.5	6255	6 AX345863	AX345863 Sequence
C 44	51	8.4	12507	6 AX345200	AX345200 Sequence
C 45	51	8.4	12578	6 AX251413	AX251413 Sequence

ALIGNMENTS

RESULT 1:  
AC134173/c  
LOCUS AC134173  
DEFINITION Bos taurus clone RP42-254113, WORKING DRAFT SEQUENCE, 26 unordered  
pieces.  
ACCESSION AC134173  
VERSION AC134173.1 GI:23306007  
KEYWORDS HTG, HTGS PHASE1, HTGS\_DRAFT.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 195534)  
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayele, M., Banks, T.,  
Buckria, J., Benton, J., Bimege, K., Blankenburg, K., Bonant, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Devila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Dolaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dith, H.H.,  
Doutwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Einhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabriel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, J.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Hollins, B.,  
Homsl, F., Howard, S., Huber, J., Hui, J., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,  
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louesged, H.,  
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,  
Moser, G., Miner, Z., Mitchell, T., Mohabat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okunolu, G.,  
Oragunye, N., Oyiedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Schere, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
Stutson, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Teifrod, B., Thomas, N., Thomas, S.,  
Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Wainstock, G., and Gibbs, R.

Direct Submission  
Unpublished.  
2 (bases 1 to 195534)  
Worley, K.C.  
Direct Submission  
Submitted (24-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 195534)  
Worley, K.C.  
Direct Submission  
Submitted (27-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
Project Information  
Center project name: VUAE  
Center clone name: RP42-254113

Summary Statistics  
Sequencing vector: M13  
Chemistry: Dye-Primer Body: 82% of reads  
Chemistry: Dye-terminator Big Dye: 18% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 185623 bases at least Q40  
Consensus quality: 190257 bases at least Q30  
Consensus quality: 192466 bases at least Q20  
Estimated insert size: 204493; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

Location/Qualifiers	1..195534
1	2358: contig of 2358 bp in length
2359	2458: gap of unknown length
2459	5619: contig of 3161 bp in length
5620	5719: gap of unknown length
5720	8310: contig of 2591 bp in length
8311	8410: gap of unknown length
8411	11002: contig of 2592 bp in length
11003	11102: gap of unknown length
11103	14386: contig of 3284 bp in length
14387	14486: gap of unknown length
14487	17149: contig of 2663 bp in length
17150	17249: gap of unknown length
17250	21414: contig of 4165 bp in length
21415	21514: gap of unknown length
21515	25171: contig of 3657 bp in length
25172	30645: gap of unknown length
30646	30745: gap of unknown length
30746	36182: contig of 5437 bp in length
36183	36282: gap of unknown length
36283	41956: contig of 5674 bp in length
41957	42056: gap of unknown length
42057	48094: contig of 6038 bp in length
48095	48194: gap of unknown length
48195	53214: contig of 5020 bp in length
53215	53314: gap of unknown length
53315	58849: contig of 5535 bp in length
58850	58949: gap of unknown length
58950	66144: contig of 7195 bp in length
66145	66244: gap of unknown length
66245	73033: contig of 6789 bp in length
73034	73133: gap of unknown length
73134	80030: contig of 6897 bp in length
80031	80130: gap of unknown length
80131	87761: contig of 7631 bp in length
87762	87861: gap of unknown length
87862	95726: contig of 7865 bp in length
95727	95826: gap of unknown length
95827	105054: contig of 9228 bp in length
105055	105154: gap of unknown length
105155	116716: contig of 11562 bp in length
116717	116816: gap of unknown length
116817	128293: contig of 11477 bp in length
128294	128393: gap of unknown length
128394	138140: contig of 9747 bp in length
138141	138240: gap of unknown length
138241	150687: contig of 12447 bp in length
150688	150787: gap of unknown length
150788	166414: contig of 15667 bp in length
166415	166514: gap of unknown length
166515	195534: contig of 29020 bp in length.

Location/Qualifiers  
1..195534  
/organism="Hos tauus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="RP42-254113"

BASE COUNT 60952 a 35784 c 36143 g 60048 t 2607 others

ORIGIN  
Query Match 23.3%; Score 141; DB 2; Length 195534;  
Best Local Similarity 100.0%; Pred. No. 2.3e-17;  
Matches 141; Conservative 0; Mismatches 0; Gaps 0;

QY 1 GAAGATTTCAGTCTCTATATAGATCTCATGATGATGATTAATACAAACAAATGAA 60  
 DB 47365 GAAGATTTCAGTCTCTATATAGATCTCATGATGATGATTAATACAAACAAATGAA 47306  
 QY 61 GGATTTCAGGATTTTAAACAGACAGTCTTTTACCAAGAAATCTCTGAGCATCAG 120  
 DB 47305 GGATTTCAGGATTTTAAACAGACAGTCTTTTACCAAGAAATCTCTGAGCATCAG 47246  
 QY 121 AATTTCATCTTCATGACTGG 141  
 DB 47245 AATTTCATCTTCATGACTGG 47225

RESULT 2  
 BC017835 566 bp mRNA linear PRI 06-DEC-2001  
 LOCUS Homo sapiens, histatin 1, clone MGC:22502 IMAGE:4289874, mRNA,  
 DEFINITION complete cds.  
 ACCESSION BC017835  
 VERSION BC017835.1 GI:17389614  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 566)  
 Strausberg,R.  
 Direct Submission  
 Submitted (03-DEC-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 REMARK  
 COMMENT Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: CLOMTECH  
 CDNA Library Preparation: CLOMTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www.shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcdpaxk1.stanford.edu](mailto:mcdpaxk1.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 36 Row: a Column: 21  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 4504528.  
 location/Qualifiers  
 1. 566  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="LocustID:3346"  
 /db\_xref="taxon:9606"  
 /clone="MGC:22502 IMAGE:4289874"  
 /rname\_type="Skeletal Muscle"  
 /clone\_lib="NIH\_MGC\_81"  
 /lab\_host="DH10B"  
 /note="Vector: pDNR-LIB"  
 68..241  
 /codon\_start=1  
 /product="histatin 1"  
 /protein\_id="AAH17835.1"  
 /db\_xref="GI:17389615"  
 /translation="MKFFVALVLAALMISMISADSHKRRHHGRKFKHSHREFF  
 FYGDGSLNYLVDN"  
 BASE COUNT 192 a 102 c 84 g 188 t  
 ORIGIN  
 Query Match 12.7%; Score 77; DB 9; Length 566;  
 Best Local Similarity 59.8%; Pred. No. 7.1e-05;

Matches 150; Conservative 0; Mismatches 95; Indels 6; Gaps 1;  
 QY 120 GAATTTCATCTTTCATGACGACGCCCAATATGAAATCTTATCTTGTTCAT 179  
 DB 34 GACTCTCTCTTGAGTAAAGACCTCAGCCACTGATGAGTTTGTCTTGTAGT 93  
 QY 180 TATGCTCTCATCTCAGCCATGATTAAGATGATTCATCTGAAAGAAACGTACAGAA 239  
 DB 94 CTGGCTCTCATGATTTTCATGATTAAGGCTGATTCATGAAAGACATCATGGGTA 153  
 QY 240 ACGGAAAAACATCATAGAGATATTTTCAACAATPACAGCCATATCAAGATATCACT 299  
 DB 154 TAGAAGAAATTCAT-----GAAAGCATCATTCATCGAATTTCCATTTATAG 207  
 QY 300 AATATATCTCTCTGCTATTCATTTCTTAAATGCTCTTATGATCTACAGACATAT 359  
 DB 208 GGACTATGATCAATATATATATGACAAATGATATTCCTTATGATATGAGGATGAT 267  
 QY 360 TAGAGAGATT 370  
 DB 268 TATAGAGTTT 278

RESULT 3  
 BD071675 357 bp DNA linear PAT 27-AUG-2002  
 LOCUS Secreted expressed sequence tags (ESTs).  
 ACCESSION BD071675  
 VERSION BD071675.1 GI:22617278  
 KEYWORDS JP 2001519667-A/485.  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 1 (bases 1 to 357)  
 Jacobs,K., McCoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D.,  
 Treacy,M., Spaulding,V. and Agostino,M.J.  
 Secreted expressed sequence tags (ESTs)  
 Patent: JP 2001519667-A 485 23-OCT-2001;  
 GENETICS INSTITUTE INC  
 OS Unidentified  
 PN JP 2001519667-A/485  
 PD 23-OCT-2001  
 PF 10-APR-1998 JP 1998543069  
 PR 10-APR-1997 US 08/838821  
 PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI  
 DAVID MERBERG,  
 PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC  
 C12N15/12, C12N5/10, C07K14/47, C12Q1/68, A61K38/17 CC, Strandedness:  
 Double;  
 CC Topology: linear;  
 CC Secreted expressed sequence tags (ESTs)  
 FH Key Location/Qualifiers  
 FT source 1..357  
 FT 1..357  
 /organism="Unidentified".  
 location/Qualifiers  
 1..357  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"  
 BASE COUNT 101 a 74 c 65 g 117 t  
 ORIGIN  
 Query Match 12.5%; Score 75.4; DB 6; Length 357;  
 Best Local Similarity 59.4%; Pred. No. 0.00015;  
 Matches 149; Conservative 0; Mismatches 96; Indels 6; Gaps 1;  
 QY 120 GAATTTCATCTTTCATGACGACGCCCAATATGAAATCTTATCTTGTTCAT 179  
 DB 48 GACTCTCTCTTGAGTAAAGACCTCAGCCACTGATGAGTTTGTCTTGTAGT 107  
 QY 180 TATGCTCTCATCTCAGCCATGATTAAGATGATTCATCTGAAAGAAACGTACAGAA 239  
 DB 94 CTGGCTCTCATGATTTTCATGATTAAGGCTGATTCATGAAAGACATCATGGGTA 167

QY 240 ACGGAAAAATCATGAGGATATTTTCAACATACGACCAATATCAAGATTCACCT 239  
DB 168 TAGAGAGAAATTCAT-----GAAAGCATCATTCACATGAGAAATTCATTTATGG 221  
QY 300 AATATATCTCTCGCGATTCATTTCTTAAATGCTGCTTAGTAACTACAGACATGAT 359  
DB 222 GGACTGTGATCAAAATATATCATATGACAAATGATATCTTAGTAATCATGCGGCATGAT 281  
QY 360 TAGAGAGATTT 370  
DB 282 TATGAGAGTTT 292

RESULT 4  
AX588147  
LOCUS AX588147 438 bp DNA linear PAT 24-JAN-2003  
DEFINITION Sequence 22 from Patent WO02083898.  
ACCESSION AX588147  
VERSION AX588147.1 GI:27899822  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Bejanin,S., Tanaka,H., Dumas Milne Edwards,J.B., Jobert,S. and  
AUTHORS Giordano,J.Y.  
TITLE Full-length human cdnas encoding potentially secreted proteins  
JOURNAL Patent: WO 02083898-A 22 24-OCT-2002;  
GENSET (FR)

FEATURES  
source location/Qualifiers  
1..438  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
84..320  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD61372.1"  
/db\_xref="GI:27899823"  
/translation="MKFVFPALVLAIMISADSHKRRHGYRRKFHEKHSYHITL  
LPLEFESSKSNANEKYNLLYTLCFRIASFIVT"  
84..140  
/note="Von Heijne matrix score 7.64030745849671 seq  
ALVLAIMISMISA/DS"

sig\_peptide  
polyA\_signal 397..402  
polyA\_site 421..438  
BASE COUNT 153 a 60 g 143 t 2 others  
ORIGIN

Query Match 11.7%; Score 70.4; DB 6; Length 438;  
Best Local Similarity 69.9%; Pred. No. 0.0013;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGAGTGCACCAATATGAGATCTTATCTTTGCTTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGACTCAGCCACATGAAATTTTGTCTTGGTTAGT 109  
QY 180 TATGCTCTCATCTTACCATGATTAAGCTGATTCATCTGAGAGAAAGCTCAGAGAA 239  
DB 110 CTGGCTCTCATGATTCATGATTAAGCTGATTCATGAGAGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAATCATCAT 255  
DB 170 TAGAGAGAAATTCAT 185

RESULT 5  
AX588483  
LOCUS AX588483 438 bp DNA linear PAT 24-JAN-2003  
DEFINITION Sequence 358 from Patent WO02083898.  
ACCESSION AX588483

VERSION AX588483.1 GI:27900156  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE Bejanin,S., Tanaka,H., Dumas Milne Edwards,J.B., Jobert,S. and  
AUTHORS Giordano,J.Y.  
TITLE Full-length human cdnas encoding potentially secreted proteins  
JOURNAL Patent: WO 02083898-A 358 24-OCT-2002;  
GENSET (FR)

FEATURES  
source location/Qualifiers  
1..438  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
84..320  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD61539.1"  
/db\_xref="GI:27900157"  
/translation="MKFVFPALVLAIMISADSHKRRHGYRRKFHEKHSYHITL  
LPLEFESSKSNANEKYNLLYTLCFRIASFIVT"  
84..140  
/note="Von Heijne matrix score 7.59999990463257 seq  
ALVLAIMISMISA/DS"

sig\_peptide  
BASE COUNT 151 a 80 c 60 g 145 t 2 others  
ORIGIN

Query Match 11.7%; Score 70.4; DB 6; Length 438;  
Best Local Similarity 69.9%; Pred. No. 0.0013;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGAGTGCACCAATATGAGATCTTATCTTTGCTTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGACTCAGCCACATGAAATTTTGTCTTGGTTAGT 109  
QY 180 TATGCTCTCATCTTACCATGATTAAGCTGATTCATCTGAGAGAAAGCTCAGAGAA 239  
DB 110 CTGGCTCTCATGATTCATGATTAAGCTGATTCATGAGAGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAATCATCAT 255  
DB 170 TAGAGAGAAATTCAT 185

RESULT 6  
AX616321  
LOCUS AX616321 438 bp DNA linear PAT 20-FEB-2003  
DEFINITION Sequence 5 from Patent WO02094864.  
ACCESSION AX616321  
VERSION AX616321.1 GI:28447365  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Bejanin,S. and Tanaka,H.  
AUTHORS Human cdnas and proteins and uses thereof  
TITLE Patent: WO 02094864-A 5 28-NOV-2002;  
JOURNAL GENSET (FR)

FEATURES  
source location/Qualifiers  
1..438  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
1..83  
84..320  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD67625.1"

```

/db_xref="GI:28447366"
/translation="MKFFVFAVLALMTSMISADSHKRRHHGKRRFHEKHSYHTL
LPLEPSSKSNANKEHYNYLTYLCRIIAFSIVT"
3'UTR
polyA_signal 318..438
polyA_site 397..402
BASE COUNT 423..438
ORIGIN 155 a 80 c 60 g 143 t

Query Match 11.7% Score 70.4; DB 6; Length 438;
Best Local Similarity 69.9%; Pred. No. 0.0013;
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCAGTCACTGAGCTCCACCAAAATATGAAGATCTTTATCTTGTTCAT 179
DB 50 GACTCTCTCTGAGAAAGAGCTCAGCAACTATGAAGTTTGTCTTGTCTTGT 109
QY 180 TATGCTCTCATCTTACGATCATATTAAGAGCTGATTCATCTGAAGAAAGCTCAGGAA 239
DB 110 CTGCTCTCTCATGATTCATGATTTATGAGCTGATTCATCATGAAAGACATCATGGTA 169
QY 240 ACGGAAAAACATCAT 255
DB 170 TAGAAGAAATTCAT 185

RESULT 7
LOCUS HUMHIS2X 524 bp mRNA linear PRI 31-DEC-1994
DEFINITION Human histatin 2 (HIS2) mRNA, complete cds.
ACCESSION M26665
VERSION M26665.1 GI:292145
KEYWORDS histatin 2.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 524)
Sabatini L.M. and Azen, B.A.
Histatins, a family of salivary histidine-rich proteins, are
encoded by at least two loci (HIS1 and HIS2)
JOURNAL Biochem. Res. Commun. 160 (2), 495-502 (1989)
MEDLINE 89246491
PUBMED 2719677
COMMENT Original source text: Homo sapiens parotid gland cDNA to mRNA.
FEATURES
source location/Qualifiers
1..524
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/ligase_type="parotid gland"
1..524
/gene="HIS2"
2..227
/gene="HIS2"
/codon_start=1
/product="histatin 2"
/protein_id="AAAS8646.1"
/db_xref="GI:292146"
/translation="MKFFVFAVLALMTSMGADSHAKRHGKRRFHEKHSRGRY
SNLYLDN"
polyA_signal 506..511
/gene="HIS2"
BASE COUNT 157 a 100 c 87 g 180 t
ORIGIN

Query Match 11.3% Score 68.2; DB 9; Length 524;
Best Local Similarity 65.4%; Pred. No. 0.0033;
Matches 100; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCAGTCACTGAGCTCCACCAAAATATGAAGATCTTTATCTTGTTCAT 179
DB 38 GATTCCTCTTGAATAAAGAGCTCAGCAACTATGAAGTTTGTGTTTGTCTTAT 97

```

```

QY 180 TATGCTCTCATCTTACGATCATATTAAGAGCTGATTCATCTGAAGAAAGCTCAGGAA 239
DB 98 CTGCTCTCATGCTTTCATGATGATGAGCTGATTCATCATGAAAGACATCATGGTA 157
QY 240 ACGGAAAAACATCATATGAGATTTTCAACA 272
DB 158 TAAAGAAATTCATGAAAGACATCATTCACA 190

RESULT 8
LOCUS BC009791 558 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, histatin 3, clone MGC:13578 IMAGE:4293405, mRNA,
complete cds.
ACCESSION BC009791
VERSION BC009791.1 GI:14602560
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 558)
Strausberg, R.
Direct Submission
Submitted (02-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
REMARK Contact: MGC help desk
Email: cga@bbs-remail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amandastembiology.org
Anup Madan, Rachel Dichtoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 19 Row: 0 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4557652.
FEATURES
source location/Qualifiers
1..558
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/db_xref="LOCUSID:3347"
/db_xref="taxon:9606"
/clone="MGC:13578 IMAGE:4293405"
/tissue_type="Skeletal Muscle"
/clone_id="NIH MGC_81"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
76..231
/codon_start=1
/product="histatin 3"
/protein_id="AAH09791.1"
/db_xref="GI:14602561"
/translation="MKFFVFAVLALMTSMGADSHAKRHGKRRFHEKHSRGRY
SNLYLDN"
BASE COUNT 187 a 102 c 91 g 178 t
ORIGIN

Query Match 11.3% Score 68.2; DB 9; Length 558;
Best Local Similarity 65.4%; Pred. No. 0.0033;
Matches 100; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCAGTCACTGAGCTCCACCAAAATATGAAGATCTTTATCTTGTTCAT 179
DB 11 TATGCTCTCATCTTACGATCATATTAAGAGCTGATTCATCTGAAGAAAGCTCAGGAA 239

```

Db 42 GATTCCTCTTGTAGTAAAGACATGACCACTAGTAGAGTTTGTGTTTCTTAAAT 101  
 QY 180 TATGGCTCATCTGATGCGCATGATGAGCTGATTCATGAGAGAAAGTCACAGAA 239  
 Db 102 CTGGGCTCATCTGATGCGCATGAGCTGATTCATGAGAGAAAGTCATGAGGTA 161  
 QY 240 ACGGAAAGAAACATCATGAGAGATTTTCAACA 272  
 Db 162 TAAAGAGAAATTCATGAGAAAGCATCATTCACA 194

RESULT 9  
 HUMSTTRNA 542 bp mRNA linear PRI 13-JAN-1995  
 LOCUS Human statherin mRNA, complete cds.  
 DEFINITION M18371.1 GI:338610  
 VERSION M18371.1 GI:338610  
 KEYWORDS statherin.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 542)  
 AUTHORS Dickinson,D.P., Ridall,A.L. and Levine,M.J.  
 TITLE Human submandibular gland statherin and basic histidine-rich peptide are encoded by highly abundant mRNA's derived from a common ancestral sequence  
 JOURNAL Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)  
 MEDLINE 88106506  
 PUBMED 3426601

COMMENT Original source text: Human female submandibular gland, cDNA to mRNA, clone PBRHSMSP98.2.  
 Draft entry and computer readable copy of sequence [1] kindly provided by D.P.Dickinson 21-MAR-1988.  
 Location/Qualifiers  
 source 1..542  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /map="4q11-q13"  
 gene 1..542  
 /gene="STATH"  
 CDS 57..245  
 /gene="STATH"  
 /note="statherin precursor"  
 /codon\_start=1  
 /protein\_id="AA060600.1"  
 /db\_xref="GI:338611"  
 /db\_xref="GDB:G00-120-391"  
 /translation="MKRLVFAFLALMVSMIGADSSSEKFLRIGRFGYGYQYQPV EGPLYQYQYQYTF"  
 sig\_peptide 57..113  
 /gene="STATH"  
 /note="statherin signal peptide"  
 mat\_peptide 114..242  
 /gene="STATH"  
 /product="statherin"  
 BASE COUNT 170 a 117 c 76 g 179 t  
 ORIGIN 262 bp upstream of PstI site; chromosome 4q11-q13.  
 Query Match 11.2%; Score 67.6; DB 9; Length 542;  
 Best local Similarity 61.2%; Pred. No. 0.0042;  
 Matches 14; Conservative 0; Mismatches 84; Indels 9; Gaps 2;

QY 140 GAACTCCACCAATATGAAGATCTTTATCTTTGCTTCATTTAGGCTTCATCTGACCA 199  
 Db 43 GAACCCAGCCCAATATGAAGATCTTTGCTTCATTTAGGCTTCATCTGACCA 102  
 QY 200 TGAATTAGAGTGAATTCATGAGAGAAAGTCACAGAGAAACGAAATCATCATGAG 259  
 Db 103 TGAATGAGCTGATTCATGAGAGAAATTTTGGCTAGAAATTTGGAAGATTCGTTATG 162  
 QY 260 GATA-----TTTCAACAATATACAGCATATCAAGATATCCACTAAATATCTCTG 313

Db 163 GGTATGCCCTTATACGACGATTCAGAACCAACCATATACCCACATACCAAC 222  
 QY 314 CGT---ATCCATTTCTTAAATGCTGCTTAGTAATACAGGACATATAGAGATTT 370  
 Db 223 AATACCAACAATATACCTTTTAAATATCATCATGAGTACGACGACATGATTTAGGCTT 282

RESULT 10  
 HUMSTATHA 552 bp mRNA linear PRI 13-JAN-1995  
 LOCUS Human statherin mRNA, complete cds.  
 DEFINITION M18078  
 ACCESSION M18078.1 GI:338507  
 VERSION M18078.1 GI:338507  
 KEYWORDS regulatory protein; statherin.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 552)  
 AUTHORS Sabatini,L.M., Carlock,L.R., Johnson,G.W. and Azen,E.A.  
 TITLE cDNA cloning and chromosomal localization (4q11-13) of a gene for statherin, a regulator of calcium in saliva  
 JOURNAL Am. J. Hum. Genet. 41 (6), 1048-1060 (1987)  
 MEDLINE 88074310  
 PUBMED 3502720

COMMENT Original source text: Human parotid gland, cDNA to mRNA, clone H772B.  
 Draft entry and computer-readable sequence [1] kindly submitted by L.Sabatini 19-JAN-1988.  
 Location/Qualifiers  
 source 1..552  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /map="4q11-q13"  
 gene 1..552  
 /gene="STATH"  
 CDS 73..261  
 /gene="STATH"  
 /note="statherin protein"  
 /codon\_start=1  
 /protein\_id="AA0594.1"  
 /db\_xref="GI:338508"  
 /db\_xref="GDB:G00-120-391"  
 /translation="MKRLVFAFLALMVSMIGADSSSEKFLRIGRFGYGYQYQPV EGPLYQYQYQYTF"  
 sig\_peptide 73..129  
 /gene="STATH"  
 /note="statherin signal peptide"  
 BASE COUNT 172 a 122 c 78 g 180 t  
 ORIGIN 278 bp upstream of PstI site; chromosome 4q11-q13.  
 Query Match 11.2%; Score 67.6; DB 9; Length 552;  
 Best local Similarity 61.2%; Pred. No. 0.0042;  
 Matches 14; Conservative 0; Mismatches 84; Indels 9; Gaps 2;

QY 140 GAACTCCACCAATATGAAGATCTTTATCTTTGCTTCATTTAGGCTTCATCTGACCA 199  
 Db 59 GAACCCAGCCCAATATGAAGATCTTTGCTTCATTTAGGCTTCATCTGACCA 118  
 QY 200 TGAATTAGAGTGAATTCATGAGAGAAAGTCACAGAGAAACGAAATCATCATGAG 259  
 Db 119 TGAATGAGCTGATTCATGAGAGAAATTTTGGCTAGAAATTTGGAAGATTCGTTATG 178  
 QY 260 GATA-----TTTCAACAATATACAGCATATCAAGATATCCACTAAATATCTCTG 313  
 Db 179 GGTATGCCCTTATACGACGATTCAGAACCAACCATATACCCACATACCAAC 238  
 QY 314 CGT---ATCCATTTCTTAAATGCTGCTTAGTAATACAGGACATATAGAGATTT 370  
 Db 239 AATACCAACAATATACCTTTTAAATATCATCATGAGTACGACGACATGATTTAGGCTT 298

RESULT 11  
AK092678 1584 bp mRNA linear PRI 15-JUL-2002  
LOCUS AK092678  
DEFINITION Homo sapiens cDNA FLJ35359 file, clone SALGL1000107, highly similar to CH-ROG PROTEIN.  
ACCESSION AK092678  
VERSION AK092678.1 GI:21751327  
KEYWORDS oligo cloning; file (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K., Arita,M., Musashino,K., Yunki,H., Hara,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Magatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isogai,T.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1584)  
AUTHORS Isogai,T. and Yamamoto,J.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FlJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
COMMENT E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
ECONOMY Trade and Industry of Japan; project supported by Ministry of Research Association for Biotechnology (RAB) / CDNA full insert sequencing: construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
FEATURES  
source location/Qualifiers  
1. 1584  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="SALGL1000107"  
/tissue\_type="salivary gland"  
/clone\_id="SALGL1"  
/note="Cloning vector: pME18FLJ3"  
89..997  
/note="unnamed protein product"  
/protein\_id="BAC03943.1"  
/db\_xref="GI:21751328"  
/translation="MGDSSEWKLKLPVDQCKEHLMKARLSGYEALKIFQKIKDEKSP EWSKFLGLIKKPVDSNAVVLKGLAALVVENAVAGKTGVSVGSVFNQPK AKAKELGIEICLMYEIEKGEAVOELLKGLNKNPKIIVACIETLRALSEFGSKII LKPIIKVLPFLFESREKARVDEALIVELVYRWIRDALRPLQINISVOLKEEEM VKLPTSAKPTKRLKRSQOELAKLEEQOSAGDAGGDDGDEVQIDAVELLEAVEI LSKLPKPTMTKLKRSQKREKPKMSL"  
CDS  
BASE COUNT 526 a 298 c 331 g 429 t  
ORIGIN  
Query Match 11.1%; Score 67.6; DB 9; Length 1584;  
Best Local Similarity 61.2%; Pred. No. 0.0036;  
Matches 147; Conservative 0; Mismatches 84; Indels 9; Gaps 2;

QY 260 GATA-----TTTCAACAATACAGCCCATATCAAGATATCACTAATATCTCCCTG 313  
DB 1203 GGTATGGCCCTTATCAGCCAGTTCAGAAACACCATATACCAACATACCAACAC 1262  
QY 314 CGT---ATCCATTTCTTAAATGCTGCTTGTACTACAGACATGATTGAGAGATT 370  
DB 1263 AATACCAACAATATACCTTTAATATATCATCATGATACAGACATGATTATGAGCTT 1322  
RESULT 12  
BD071749 203 bp DNA linear PAT 27-AUG-2002  
LOCUS BD071749  
DEFINITION Secreted expressed sequence tags (sests).  
ACCESSION BD071749  
VERSION BD071749.1 GI:22617352  
KEYWORDS JP 2001519667-A/559.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 203)  
AUTHORS Jacobs,K., McCoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D., Treacy,M., Spaulding,V. and Agostino,M.J.  
TITLE Secreted expressed sequence tags (sests)  
JOURNAL Patent: JP 2001519667-A 559 23-OCT-2001;  
COMMENT GENETICS INSTITUTE INC  
OS Unidentified  
PN JP 2001519667-A/559  
PD 23-OCT-2001  
PF 10-APR-1998 JP 1998543069  
PR 10-APR-1997 US 08/838821  
PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI DAVID MERBERG,  
PI MAURICE TREACY, VIKKI SPULDING, MICHAEL J AGOSTINO PC  
C1N15/12.C1N15/10.C07K14/47.C1201/68.A61K38/17 CC Strandedness:  
Double;  
CC Topology: Linear;  
CC Secreted expressed sequence tags (sests)  
FH Key location/Qualifiers  
FT source 1.203  
/organism="Unidentified".  
FEATURES  
source location/Qualifiers  
1. 203  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 59 a 43 c 39 g 62 t  
ORIGIN  
Query Match 11.1%; Score 67.2; DB 6; Length 203;  
Best Local Similarity 68.4%; Pred. No. 0.0059;  
Matches 93; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 120 GAATTCATCTTTTCATGACTGACCAACAATATGAAATCTTTATCTTTGCTTCAT 179  
DB 47 GATTCCTCTTGAATGAAAGACATGACCAATATGAAATTTTGTGTTTGTCTTAT 106  
QY 180 TATGCTCTCATCTAGCAATGATTAAGCTGATTCATCTGAAGAAGCTCACAGAA 239  
DB 107 CTGGCTCTCATGCTTTCCATGACTGAGCTGATTCATCATGCAAGAAGACATCATGGGTA 166  
QY 240 ACGGAAAAACATCAT 255  
DB 167 TAAAGAAATTCAT 182  
RESULT 13  
HMHHSIX 480 bp mRNA linear PRI 31-DEC-1994  
LOCUS HMHHSIX  
DEFINITION Human histatin 1 (HIS1) mRNA, complete cds.  
ACCESSION M26664  
VERSION M26664.1 GI:292143  
KEYWORDS histatin 1.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 480)  
AUTHORS Sabatini, L.M. and Azen, E.A.  
TITLE Histatins, a family of salivary histidine-rich proteins, are  
encoded by at least two loci (HIS1 and HIS2)  
JOURNAL Biochem. Biophys. Res. Commun. 160 (2), 495-502 (1989)  
MEDLINE 89246491  
PUBMED 2719677  
COMMENT Original source text: Homo sapiens parotid gland cDNA to mRNA.  
FEATURES  
source  
Location/Qualifiers  
1..480  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="parotid gland"  
1..480  
/gene="HIS1"  
10..183  
/codon\_start=1  
/product="histatin 1"  
/protein\_id="AA58645.1"  
/db\_xref="GI:292144"  
/translation="MKPFVFLVLMISMISADSHKRRHGYRKFHEKHSREPP  
FYDYGSNLYDN"  
polya\_signal 462..467  
/gene="HIS1"  
BASE COUNT 152 a 85 c 75 g 168 t  
ORIGIN  
Query Match 11.0%; Score 66.6; DB 9; Length 480;  
Best Local Similarity 59.6%; Pred. No. 0.0067;  
Matches 133; Conservative 0; Mismatches 84; Indels 6; Gaps 1;  
QY 148 CCAAAATGAGATCTTATCTTGTCTTCAATATGCTCATCTAGCCATGATTAGA 207  
Db 4 CCAACTATGAGATCTTATCTTGTCTTCAATATGCTCATCTAGCTCATGATTAGC 63  
QY 208 GCTGATTCATCTGAGAGAAAGCGTCACAGAGAAAGGAAAAACATCTAGAGATATTTT 267  
Db 64 GCTGATTCATCTGAGAGAAAGCGTCACAGAGATGCTATGAGAAATTCAT-----GAAAG 117  
QY 268 CAACAATACGACGATATCAACGATATCACTAATATATCTCTGCGTATCCATTTCT 327  
Db 118 CATCATTCACATCGAGATTTCCATTTTATGCGGACATATGATCAATTTATCTATATGAC 177  
QY 328 TAAAAATGCTGCTTAGTAACTACAGACATGATTAGAGATTT 370  
Db 178 AATTGATATCTTATGATATATGAGGCGATGATTATAGAGGTTT 220  
RESULT 14  
AY154893 486 bp mRNA linear MAM 17-DEC-2002  
LOCUS Bos taurus statherin (STATH) mRNA, complete cds.  
DEFINITION AY154893  
ACCESSION AY154893  
VERSION AY154893.1 GI:27227454  
KEYWORDS Bos taurus (cow)  
SOURCE Bos taurus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Rajnkeles, M., Elmitcki, L., Miller, W. and Rosen, J.M.  
TITLE Multi-species comparative sequence analysis of the casein gene  
cluster region: A gene domain encoding epithelial secretory  
proteins  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 486)  
AUTHORS Rajnkeles, M. and Rosen, J.M.

TITLE Direct Submission  
JOURNAL Submitted (24-SEP-2002) Molecular and Cellular Biology, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
FEATURES  
source  
Location/Qualifiers  
1..486  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/chromosome="6"  
/tissue\_type="salivary gland"  
1..486  
/gene="STATH"  
1..180  
/codon\_start=1  
/product="statherin"  
/protein\_id="AA85581.1"  
/db\_xref="GI:27227455"  
/translation="MKIFFFAFLMALVMIKADSSSEEHRLFNRPFYYPNOGGYI  
PSYPAYPYPPYVQ"  
BASE COUNT 155 a 92 c 79 g 160 t  
ORIGIN  
Query Match 10.9%; Score 65.6; DB 4; Length 486;  
Best Local Similarity 62.8%; Pred. No. 0.01;  
Matches 123; Conservative 0; Mismatches 64; Indels 9; Gaps 1;  
QY 154 ATGAAGATCTTTATCTTGTCTTCATTATGCGCTCATCTAGCCATGATTAGCGTAT 213  
Db 1 ATGAAGATCTTTATCTTGTCTTCATTATGCGCTCATCTAGCCATGATTAGCGTAT 60  
QY 214 TCATCTGAGAGAAAGCTCACAGAAACGAAAAACATCATAGAGATATTTTCAACAA 273  
Db 61 TCATCTGAGAGAGAAC-----ACCGCTTAGATTTAACCTAGATTATTATACCA 111  
QY 274 TACCAGCCATATCAAGATATCCATTAATTAATTCCTCGATCCATTTCTTAAAT 333  
Db 112 AACCGAGAGGTGATATATCTTCTAGTTATCCAGATATCATATTCATATCAATCA 171  
QY 334 GCTGCTTAGTAATAC 349  
Db 172 GTCCAGTAGTACTGC 187  
RESULT 15  
HUMBRPA 491 bp mRNA linear PRI 31-OCT-1994  
LOCUS Human histidine-rich protein mRNA, complete cds.  
DEFINITION M18372  
ACCESSION M18372.1 GI:179465  
VERSION M18372.1  
KEYWORDS histidine-rich protein.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 491)  
AUTHORS Dickinson, D.P., Ridall, A.L. and Levine, M.J.  
TITLE Human submandibular gland statherin and basic histidine-rich  
peptide are encoded by highly abundant mRNA's derived from a common  
ancestral sequence  
JOURNAL Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)  
MEDLINE 88106506  
PUBMED 3426601  
COMMENT Original source text: Human female submandibular gland, cDNA to  
mRNA, clone pBRHNSFI8.1.  
Draft entry and computer readable copy of sequence [1] kindly  
provided by D.P. Dickinson 21-MAR-1988.  
FEATURES  
source  
Location/Qualifiers  
1..491  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="3q21-quer"





**THIS PAGE BLANK (USPTO)**

12.5	357	20	AAV89526	EST clone
------	-----	----	----------	-----------

Bovine histatin II  
EST clone CP294.

9	70.4	11.7	438	25	ACC51062	Human Chimerin codon
10	70.4	11.7	438	25	ABZ35425	Human GENSST codin
11	70.4	11.7	438	25	ABZ35590	Human GENSST codin
12	68.2	11.3	857	22	AAH96586	Human EST-derived
13	67.6	11.2	552	21	AAA48964	Human strathern DN
14	67.2	11.1	203	20	AAV86900	EST clone CP41. H
15	62.6	10.4	491	21	AAA48965	Human basic histidi
16	60.8	10.1	516	20	AAV89525	EST clone CP289.
17	58	9.6	457	20	AEQ95552	Human coding sequen
18	55.8	9.2	176	20	AAV89602	EST clone CP92. H
C 19	53.6	8.9	13449	29	ABL33385	Human immune syste
C 20	51.2	8.5	6255	24	ABL33961	Human immune syste
C 21	51	8.4	12507	24	ABL32298	Human immune syste
C 22	51	8.4	12578	22	AA546659	Human immune syste
C 23	49.8	8.2	7106	22	ABL33478	Human immune syste
C 24	49.4	8.2	5945	24	ABL33084	Human immune syste
C 25	49.4	8.2	16994	24	ABL32989	Human immune syste
C 26	49.2	8.1	13131	24	ABL92742	Chemically treated
C 27	49	8.1	73334	24	ABL92319	Chemically treated
C 28	49	8.1	73334	24	ABL34125	Human immune syste
C 29	47.8	7.9	6782	24	AA543777	Human immune syste
C 30	47.6	7.9	6274	22	AA546321	Tumour suppressor
C 31	47.6	7.9	6674	24	ABK31238	Signal transductio
C 32	47.6	7.9	19659	24	ABL33765	Human immune syste
C 33	47.2	7.8	6155	24	ABK28362	DNA transcription
C 34	47.2	7.8	6280	24	ABL92425	Chemically treated
C 35	47	7.8	11691	24	ABL34241	Human immune syste
C 36	47	7.8	15732	22	AA545388	Human immune syste
C 37	47	7.8	15732	22	ABK28233	Chemically pretrea
C 38	46.6	7.7	5364	24	ABK40005	DNA transcription
C 39	46.4	7.7	6311	24	ABL33963	Human chemically p
C 40	46.4	7.7	7187	24	ABN80036	Human immune syste
C 41	46.2	7.6	5641	24	ABL33396	Human chemically m
C 42	46.2	7.6	12138	24	ABL33943	Human immune syste
C 43	46.2	7.6	12138	24	ABK28336	DNA transcription
C 44	46.2	7.6	17738	24	ABL33539	Human immune syste
C 45	46	7.6	5689	22	AA545384	Chemically pretrea

## ALIGNMENTS

RESULT 1  
ABX11391

AC ABX1139

DT 01-MAY-2003 (first entry)

DE Bovine histatin like polypeptide, #4, cDNA.

Cow; gene; ss; histatin; bovine; salivary secretion; oral cavity;

KW microbial infection; fungal infection; dental caries; plaque; tartar;

KW fungicide; antibacterial.

OS Bos taurus

FH	Key	Location/Qualifiers
1		

```
/*tag= a
```

[illegible]

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

PR 29-OCT-1999; 99US-162701P

PR 22-AUG-2000; 2000US-0644190.  
 PR 27-OCT-2000; 2000US-0699146.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 PI Glenn M, Grigor MR, Molenaar AJ, Davis SR;  
 XX WPI; 2003-275306/27.  
 DR P-PSDB; ABG75794.  
 XX  
 PT New histatin polynucleotides and polypeptides expressed in bovine  
 PT tissues, useful for treating bacterial or fungal infections in mammals,  
 PT e.g. infections of the oral cavity, vagina, urethra, ear or skin, or  
 PT systemic infections -  
 XX  
 PS Claim 1; Page 12; 17pp; English.  
 XX  
 CC The invention discloses isolated histatin polynucleotides and  
 CC polypeptides, which comprise sequences expressed in bovine. Histatins are  
 CC histidine rich and specific to the salivary secretions. They are believed  
 CC to function as part of the non-immune defence system, particularly in the  
 CC oral cavity and have promise as therapeutic agents in humans with oral  
 CC candidosis. Also disclosed are compositions comprising the histatin  
 CC polypeptide or polynucleotide, and at least one component consisting of  
 CC physiological or pharmaceutical carriers or immunostimulants. The  
 CC histatin polynucleotide, polypeptide, or the cosmetic composition  
 CC comprising the polypeptide, is useful for treating (e.g. gene therapy) a  
 CC disorder in a mammal, particularly microbial or fungal infections. The  
 CC histatin polynucleotide is also useful in genome mapping, physical  
 CC mapping or in the positional cloning of genes. Specifically, the  
 CC polypeptide or polynucleotide is useful for treating fungal or bacterial  
 CC infections of the oral cavity (e.g. dental caries, plaque or tartar),  
 CC vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis  
 CC caused by *Pseudomonas aeruginosa*), mucosa or eye, as well as for treating  
 CC systemic infections such as systemic *Candida* infection. The polypeptide  
 CC or polynucleotide is also useful for veterinary applications e.g. for  
 CC treating mastitis. The sequence presented is a cDNA encoding a bovine  
 CC histatin like polypeptide.  
 CC  
 XX  
 SQ Sequence 604 BP; 222 A; 112 C; 79 G; 191 T; 0 other;

Query Match 100.0%; Score 604; DB 25; Length 604;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-120;  
 Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTATTTTCACTTCTATATATAGATCTCATATCTGATGTAATTAACAAACAAAGAA 60  
 DB 1 GAAGTATTTTCACTTCTATATATAGATCTCATATCTGATGTAATTAACAAACAAAGAA 60  
 QY 61 GGATTTCAAGGTATTTAAACACAGAGTTTCTGCAAGAAACATCTCTGAAGCATCAG 120  
 DB 61 GGATTTCAAGGTATTTAAACACAGAGTTTCTGCAAGAAACATCTCTGAAGCATCAG 120  
 QY 121 AATTTCATCTTTCATGACGTGACTCCACCAATATGAAAGATCTTTATCTTTGTTCTTCA 180  
 DB 121 AATTTCATCTTTCATGACGTGACTCCACCAATATGAAAGATCTTTATCTTTGTTCTTCA 180  
 QY 181 ATGGCTCTCATCTCAGCAGATTTAGAGCTGATTCATCTGAAGAAACGTCACAGAAA 240  
 DB 181 ATGGCTCTCATCTCAGCAGATTTAGAGCTGATTCATCTGAAGAAACGTCACAGAAA 240  
 QY 241 CGGAAAAAATCATATAGAGATTTTCAACAAATACAGAGCATATCAAGATATCCACTA 300  
 DB 241 CGGAAAAAATCATATAGAGATTTTCAACAAATACAGAGCATATCAAGATATCCACTA 300  
 QY 301 AATTATCTCTCTGCGATTCATTTCTTAATAATGCTGTTAGTACATACAGACATGAT 360  
 DB 301 AATTATCTCTCTGCGATTCATTTCTTAATAATGCTGTTAGTACATACAGACATGAT 360  
 QY 361 AGAGAGATTTTTCACATGATTTTCTACTCTTTCTGTTGTTGAAAAACATCTTTCA 420  
 DB 361 AGAGAGATTTTTCACATGATTTTCTACTCTTTCTGTTGTTGAAAAACATCTTTCA 420

QY 421 AATGAATAAAAACAAAGAAAAAATATCAGTCAAGTATGTCACACATACATCTGGATC 480  
 DB 421 AATGAATAAAAACAAAGAAAAAATATCAGTCAAGTATGTCACACATACATCTGGATC 480  
 QY 481 AATATCAATATTTTAAAAATATATATGATGATCTCTGACATATGATTAATGGTTCTAC 540  
 DB 481 AATATCAATATTTTAAAAATATATATGATGATCTCTGACATATGATTAATGGTTCTAC 540  
 QY 541 TTCTTTTCTCTGCTGACTTACCATGATGCTTATATATGATCTATCAAGCATAAAAA 600  
 DB 541 TTCTTTTCTCTGCTGACTTACCATGATGCTTATATATGATCTATCAAGCATAAAAA 600  
 QY 601 AAAA 604  
 DB 601 AAAA 604

RESULT 2  
 ABX11388  
 ID ABX11388 standard; cDNA; 505 BP.  
 XX  
 AC ABX11388;  
 XX  
 DT 01-MAY-2003 (first entry)  
 XX  
 DE Bovine histatin like polypeptide, #1, cDNA.  
 XX  
 KW Cow; gene; ss; histatin; bovine; salivary secretion; oral cavity;  
 KW non-immune defence system; oral candidosis; gene therapy;  
 KW microbial infection; fungal infection; dental caries; plaque; tartar;  
 KW cystic fibrosis; systemic infection; *Candida* infection; mastitis;  
 KW fungicide; antibacterial.  
 XX  
 OS Bos taurus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 44..220  
 FT /\*tag= a  
 FT /product= "Histatin like polypeptide"  
 XX  
 PN US2002164625-A1.  
 XX  
 PD 07-NOV-2002.  
 XX  
 PF 19-FEB-2002; 2002US-0079754.  
 XX  
 PR 23-AUG-1999; 99US-150330P.  
 PR 29-OCT-1999; 99US-162701P.  
 PR 22-AUG-2000; 2000US-0644190.  
 PR 27-OCT-2000; 2000US-0699146.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 PI Glenn M, Grigor MR, Molenaar AJ, Davis SR;  
 XX WPI; 2003-275306/27.  
 DR P-PSDB; ABG75791.  
 XX  
 PT New histatin polynucleotides and polypeptides expressed in bovine  
 PT tissues, useful for treating bacterial or fungal infections in mammals,  
 PT e.g. infections of the oral cavity, vagina, urethra, ear or skin, or  
 PT systemic infections -  
 XX  
 PS Claim 1; Page 11; 17pp; English.  
 XX  
 CC The invention discloses isolated histatin polynucleotides and  
 CC polypeptides, which comprise sequences expressed in bovine. Histatins are  
 CC histidine rich and specific to the salivary secretions. They are believed  
 CC to function as part of the non-immune defence system, particularly in the  
 CC oral cavity and have promise as therapeutic agents in humans with oral  
 CC candidosis. Also disclosed are compositions comprising the histatin  
 CC polypeptide or polynucleotide, and at least one component consisting of  
 CC physiological or pharmaceutical carriers or immunostimulants. The



```

DB 373 TGAATTTCTTCTACTCTTCTTCTGTGTGAAACATCTTCAATGAATGAACAAAGA 432
QY 438 AAAAAAAAAAGTAACTAGTGTGACAAACATCTTGAATGAATGAATGAATTTTAA 497
DB 433 AAAAAAAAAAGTAACTAGTGTGACAAACATCTTGAATGAATGAATGAATTTTAA 492
QY 498 AACATAATTAATGAATGCTGTGAACTATGTAATGTTCTTCTTCTCTGTAC 557
DB 493 AACATAATTAATGAATGCTGTGAACTATGTAATGTTCTTCTTCTCTGTAC 552
QY 558 TTACCATGATGCTTAAATTAATGATCTATCA 590
DB 553 TTACCATGATGCTTAAATTAATGATCTATCA 585

RESULT 4
ABX11393
ID ABX11393 standard; cDNA; 525 BP.
XX
XX ABX11393;
XX
XX 01-MAY-2003 (first entry)
XX
XX Bovine histatin like polypeptide, #6, cDNA.
XX
XX Cow; gene; ss; histatin; bovine; salivary secretion; oral cavity;
XX non-immune defence system; oral candidosis; gene therapy;
XX microbial infection; fungal infection; dental caries; plaque; tartar;
XX cystic fibrosis; systemic infection; Candida infection; mastitis;
XX fungicide; antibacterial.
XX
XX Bos taurus.
XX
XX
XX
XX
XX Key Location/Qualifiers
XX CDS 43..222
XX /*tag= a
XX /product= "Histatin like polypeptide"
XX
XX
XX US2002164625-A1.
XX
XX
XX 07-NOV-2002.
XX
XX
XX 19-FEB-2002; 2002US-0079754.
XX
XX
XX 23-AUG-1999; 99US-150330P.
XX 29-OCT-1999; 99US-162701P.
XX 22-AUG-2000; 2000US-064150.
XX 27-OCT-2000; 2000US-0699146.
XX
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX
XX Glenn M, Grigor MR, Molenaar AJ, Davis SR;
XX
XX WPI; 2003-275306/27.
XX P-PSDB; ABG75796.
XX
XX
XX New histatin polynucleotides and polypeptides expressed in bovine
XX tissues, useful for treating bacterial or fungal infections in mammals,
XX e.g. infections of the oral cavity, vagina, urethra, ear or skin, or
XX systemic infections
XX
XX
XX Claim 1; Page 13; 17pp; English.
XX
XX
XX The invention discloses isolated histatin polynucleotides and
XX polypeptides, which comprise sequences expressed in bovine. Histatins are
XX histidine rich and specific to the salivary secretions. They are believed
XX to function as part of the non-immune defence system, particularly in the
XX oral cavity and have promise as therapeutic agents in humans with oral
XX candidosis. Also disclosed are compositions comprising the histatin
XX polypeptide or polynucleotide, and at least one component consisting of
XX physiological or pharmaceutical carriers or immunostimulants. The
XX histatin polynucleotide, polypeptide, or the cosmetic composition

```

```

CC comprising the polypeptide, is useful for treating (e.g. gene therapy) a
CC disorder in a mammal, particularly microbial or fungal infections. The
CC histatin polynucleotide is also useful in genome mapping, physical
CC mapping or in the positional cloning of genes. Specifically, the
CC polypeptide or polynucleotide is useful for treating fungal or bacterial
CC infections of the oral cavity (e.g. dental caries, plaque or tartar),
CC vagina, urethra, ear, skin, respiratory tract (e.g. cystic fibrosis
CC caused by Pseudomonas aeruginosa), mucosa or eye, as well as for treating
CC systemic infections such as systemic Candida infection. The polypeptide
CC or polynucleotide is also useful for veterinary applications e.g. for
CC treating mastitis. The sequence presented is a cDNA encoding a bovine
CC histatin like polypeptide.
XX
XX
XX Sequence 525 BP; 187 A; 102 C; 72 G; 164 T; 0 other;
XX
XX
XX Query Match 70.1%; Score 423.4; DB 25; Length 525;
XX Best Local Similarity 93.4%; Pred. No. 7.7e-82;
XX Matches 466; Conservative 0; Mismatches 1; Indels 32; Gaps 1;
XX
XX
XX QY 138 CTGAGCTCCACCAATATGATGATCTTATCTTGTCTTCAATATGCTTCATCTAGC 197
XX 27 CAGAGCTCCACCAATATGATGATCTTATCTTGTCTTCAATATGCTTCATCTAGC 86
DB
XX
XX QY 198 CATGATTAGAGCTGATTCATCTGAAGAAACCTCACAGAAACGAAAAACATCAT-- 255
XX 87 CATGATTAGAGCTGATTCATCTGAAGAAACCTCACAGAAACGAAAAACATCATGT 146
DB
XX
XX QY 256 -----AGAGATATTTTCAAGATACAGCATAT 285
XX 147 TGAATAGCTCCAGAAATCTTACTAATPACAGAGAAATTTTCAACATACAGCATAT 206
DB
XX
XX QY 286 CAACGATATCACTAATATATCTCTGCTATCATCTTCTTAAATGCTGCTAGTA 345
XX 207 CAACGATATCACTAATATATCTCTGCTATCATCTTCTTAAATGCTGCTAGTA 266
DB
XX
XX QY 346 CTACAGGACATGATTAAGAGATTTTCAACATGATTTTCTTCTTCTTCTTGT 405
XX 267 CTACAGGACATGATTAAGAGATTTTCAACATGATTTTCTTCTTCTTCTTGT 326
DB
XX
XX QY 406 GAAACCATCTTCAANTGAATGAATGAACAAAGAAAAAATCGTCAAGATTGACAA 465
XX 327 GAAACCATCTTCAANTGAATGAATGAACAAAGAAAAAATCGTCAAGATTGACAA 386
DB
XX
XX QY 466 CACATCTTGAATCAATATCAATATTTTAAACATATATGATAGTCTGAACTAT 525
XX 387 CACATCTTGAATCAATATCAATATTTTAAACATATATGATAGTCTGAACTAT 446
DB
XX
XX QY 526 GTAATTGTTTCTTCTTCTTCTTCTGTCATTCATGATGATGATTAATGATCT 585
XX 447 GTAATTGTTTCTTCTTCTTCTTCTGTCATTCATGATGATGATTAATGATCT 506
DB
XX
XX QY 586 ATCAAGCATTAATAAAAAA 604
XX 507 ATCAAGCATTAATAAAAAA 525
DB

RESULT 5
ABX11394
ID ABX11394 standard; cDNA; 267 BP.
XX
XX
XX ABX11394;
XX
XX
XX 01-MAY-2003 (first entry)
XX
XX
XX Bovine mammary tissue cDNA clone.
XX
XX
XX Cow; ss; histatin; bovine; salivary secretion; oral cavity;
XX non-immune defence system; oral candidosis; gene therapy;
XX microbial infection; fungal infection; dental caries; plaque; tartar;
XX cystic fibrosis; systemic infection; Candida infection; mastitis;
XX fungicide; antibacterial; mammary-gland.
XX
XX Bos taurus.

```



cc. histatin polynucleotide, polypeptide, or the cosmetic composition

CC stimulating or suppressing activity, haematopoiesis regulating



CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
XX  
SQ Sequence 357 BP; 101 A; 74 C; 65 G; 117 T; 0 other;

Query Match 12.5%; Score 75.4; DB 20; Length 357;  
Best Local Similarity 59.4%; Pred. No. 4.3e-07;  
Matches 149; Conservative 0; Mismatches 96; Indels 6; Gaps 1;

QY 120 GAATTCATCTTTCATGACTGACCTCCACCAATATGAGATCTTATCTTGTCTTCAT 179  
DB 48 GACTCTCTCTTGAGTAAAGACTCAGCCACTATGAAATTTTGTCTTGTCTTGTAGT 107  
QY 180 TATGGCTCTCATCCATGACCATGATTTAGACTGATTCATTTGAAGAAACGTCACAGAA 239  
DB 108 CTGGCTCTCATGATTTCCATGATTTAGCGCTATTCACATGAAAGACATCATGGGTA 167  
QY 240 ACGGAAAAACATCATAGAGATATTTTCACACATACCGCATATTCACGATTCACACT 299  
DB 168 TAGAAGAAAATTCAT-----GAAAGCATCATTCACATCGAATTTCCATTTTANGG 221  
QY 300 AATTTATCTCTCGGTATCCATTTCTTAAATGCTGCTTGAATCTACAGACATGAT 359  
DB 222 GGACTGTGGATCAATTAATCTATGACATTTGATATCTTATGATATGAGGATGAT 281  
QY 360 TAGAGAGATTT 370  
DB 282 TATAGAGGTTT 292

## RESULT 9

ACC51062  
ID ACC51062 standard; cDNA; 438 BP.

AC ACC51062;

DT 13-JUN-2003 (first entry)

DE Human Chimerin coding sequence.

KW Human; GENSET; therapeutic; therapy; gene; ss.

OS Homo sapiens.

PN WO200294864-A2.

PD 28-NOV-2002.

PF 06-AUG-2001; 2001WO-1B01715.

PR 25-MAY-2001; 2001US-293574P.

PR 15-JUN-2001; 2001US-298698P.

PR 29-JUN-2001; 2001US-302277P.

PR 13-JUL-2001; 2001US-305456P.

PA (GENSET) GENSET.

PI Benjamin S, Tanaka H;

DR WPI: 2003-129412/12.

DR P-PsDB; ABR48455.

XX New GENSET polynucleotides and polypeptides, useful for preparing a  
PT composition for treating GENSET-related disorders and as reagents in  
PT assays to quantitatively determined levels of GENSET expression in  
PT biological samples -  
XX  
PS Claim 1; Page 404-405; 505pp; English.

CC The present invention relates to novel human GENSET coding sequences

CC (ACC51060-ACC51115) and proteins (ABR48453-ABR48508). The GENSET  
CC sequences are useful for preparing a composition for treating  
CC GENSET-related disorders. They can also be used as markers for tissues in  
CC which the corresponding protein is preferentially expressed, as molecular  
CC weight markers on Southern gels, as chromosome markers or tags to  
CC identify chromosomes, and as reagents in assays to quantitatively  
CC determined levels of GENSET expression in biological samples.  
XX  
SQ Sequence 438 BP; 155 A; 80 C; 60 G; 143 T; 0 other;

Query Match 11.7%; Score 70.4; DB 25; Length 438;  
Best Local Similarity 69.9%; Pred. No. 5.4e-06;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACTGACCTCCACCAATATGAGATCTTATCTTGTCTTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGACTCAGCCACTATGAAATTTTGTCTTGTCTTGTAGT 109  
QY 180 TATGGCTCTCATCCATGACCATGATTTAGAGCTGATTCATCTGAAGAAACGTCACAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTTAGCGCTGATTCATGATGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAAGAAAATTCAT 185

## RESULT 10

ABZ36425  
ID ABZ36425 standard; cDNA; 438 BP.

AC ABZ36425;

DT 21-FEB-2003 (first entry)

DE Human GENSET coding sequence SEQ ID 22.

KW Cytostatic; antiinflammatory; nootropic; neuroprotective; cardiant;

KW gastrointestinal; gene therapy; GENSET; heavy metal toxicity; cancer;

KW inflammatory disease; immune disorder; neuromuscular; toxicity;

OS central nervous system; cardiovascular; gastrointestinal; gene; ss.

PN Homo sapiens.

PN WO200283898-A1.

PD 24-OCT-2002.

PF 18-APR-2001; 2001WO-1B00914.

PF 18-APR-2001; 2001WO-1B00914.

PA (GENSET) GENSET.

PI Benjamin S, Tanaka H, Dumas Milne Edwards J, Robert S, Giordano J;

DR WPI: 2003-075548/07.

DR the toxicity -

XX New GENSET polynucleotides and polypeptides, useful for treating heavy  
PT metal toxicity, cancer, inflammatory diseases, immune disorders, and  
PT the neuromuscular, CNS, cardiovascular or gastrointestinal effects of  
PT the toxicity -  
XX  
PS Claim 12; Page 295; 735pp; English.

CC The present invention relates to novel GENSET polynucleotides  
CC (ABZ36404-ABZ36911) encoding polypeptides (ABP75963-ABP76368). The  
CC polynucleotides and polypeptides are useful in screening and diagnostic  
CC assays for abnormal GENSET expression and/or biological activity. They  
CC are also useful for screening of compounds for treating or preventing  
CC GENSET-related disorders, such as heavy metal toxicity, cancer,  
CC inflammatory diseases, immune disorders, and the neuromuscular, central  
CC nervous system (CNS), cardiovascular or gastrointestinal effects of the



QY 120 GAATTCATCTTTCATGACTGACGACCAAAATATGAAAGATCTTATCTTGTCTTCAT 179  
DB 361 GATTCCTCCCTTGTAGAAAGAGCTGACGACCACTATGAAATTTTGTGTTTGTAT 420  
QY 180 TATGCTCATCCATGACCATATGAGCTGATTCATGTAAGAAACGTCACAGAA 239  
DB 421 CTGGCTCTCAATGCTTTCATGATGAGCATATTCACATGCAAGAGACATCATGAGTA 480  
QY 240 ACGAAAAAATCATGAGATATTTTCAACA 272  
DB 481 TAAAGAAAAATTCATGAAAGACATCATTCACA 513

RESULT 13  
AAAA8964  
ID AAAA8964 standard; DNA; 552 BP.  
XX  
AC AAAA8964;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human statherin DNA.  
XX  
Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;  
KM precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;  
KM asthma; allergy; diabetes mellitus; fungal; bacterial infection;  
KM cancer; leukemia; adenocarcinoma; melanoma; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 73..261  
FT /tag= a  
FT /product= Statherin  
XX  
XX WO200024779-A1.  
XX  
XX 04-MAY-2000.  
XX  
XX 22-OCT-1999; 99WO-US24046.  
XX  
XX 23-OCT-1998; 98US-0155209.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX  
XX Tang YT, Corley NC, Guegler KJ, Patterson C;  
XX  
XX WPI; 2000-350699/30.  
XX  
XX P-PSDB; AAY94527.  
XX  
XX Purified polypeptide used for treating or preventing a disorder  
XX characterized by expression or activity of lysine-rich statherin  
XX protein -  
XX  
XX Disclosure; Page 71; 75pp; English.  
XX  
XX The present invention relates to human lysine-rich statherin protein  
XX (LRSP) (AAY94526). The cDNA sequence encoding this protein was identified  
XX through analysis of a cDNA library of breast tumour tissue (BRSTW014).  
XX The LRSP sequence was found to have homology with the DNA of human  
XX statherin (the present sequence) and human basic histidine-rich protein  
XX (AAY94528). Human statherin is a phosphoprotein that acts as an  
XX inhibitor of precipitation of calcium phosphate salts in the oral cavity.  
XX The LRSP polypeptide and its antagonists may be useful for treating or  
XX preventing disorders associated with the activity of LRSP. Such  
XX disorders include autoimmune/inflammatory disorders (for example AIDS,  
XX allergies, asthma, diabetes mellitus), bacterial and fungal infection  
XX and cancers (such as leukemia, adenocarcinoma, melanoma). Antibodies to  
XX LRSP may be useful for diagnosis of the above disorders.  
XX  
XX Sequence 552 BP; 172 A; 122 C; 78 G; 180 T; 0 other;  
XX  
Query Match 11.2%; Score 67.6; DB 21; Length 552;

Best Local Similarity 61.2%; Pred. No. 2.2e-05;  
Matches 147; Conservative 0; Mismatches 84; Indels 9; Gaps 2;  
QY 140 GCATCCCAACCAATATGAAAGATCTTATCTTTCCTCATATAGGCTCATCCAGCA 199  
DB 59 GAACCCAGCCCACTATGAAATTCCTTCTTGTGCTCATCTGCTCTCATGTTTCA 118  
QY 200 TGATTAGAGCTGATTCATCTGTAAGAAACGTCACAGAAACGAAAAATCATCATAG 259  
DB 119 TGATTGAGCTGATTCATCTGTAAGAAATTTTGGCGTAAATGGAAGATTGGTTATG 178  
QY 260 GATA-----TTTCAACAATACAGCCATATCAAGATATCCATAATTATCTCTG 313  
DB 179 GGTATGCCCTTATCAGCCAGTTCAGAAACCACTATACCCCAACCATACCAACAC 238  
QY 314 CGT---ATCCATTTTCCTTAAATGCTGTTAGTAACTACAGACATGATTAGAGATT 370  
DB 239 AATACCAACAATATACCTTTTAATATCATCATGTAACGAGACATGATTAATGAGCTT 298

RESULT 14  
AAV89600  
ID AAV89600 standard; cDNA; 203 BP.  
XX  
XX AAV89600;  
XX  
XX 15-FEB-1999 (first entry)  
XX  
XX EST clone CP41.  
XX  
XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
XX tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
XX receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
XX gene therapy; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO9845436-A2.  
XX  
XX 15-OCT-1998.  
XX  
XX 10-APR-1998; 98WO-US06955.  
XX  
XX 10-APR-1997; 97US-0838821.  
XX  
XX (GENY) GENETICS INST INC.  
XX  
XX Agostino MJ, Jacobs K, Lavallic ER, McCoy JM, Merberg D;  
XX Racie LA, Spaulding V, Treacy M;  
XX  
XX WPI; 1999-070077/06.  
XX  
XX New polynucleotides encoding human secreted proteins - derived from  
XX e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
XX ovary, pituitary, retina and colon cDNA libraries.  
XX  
XX Claim 1; Page 263-270; 618pp; English.  
XX  
XX The present sequence represents a human expressed sequence tag (EST).  
XX The polynucleotide, which is a secreted EST, and the encoded protein  
XX are predicted to have useful biological activities which would make  
XX them suitable for treating, preventing or ameliorating medical  
XX conditions in humans and animals, although no supporting data is  
XX given. Suggested activities include nutritional activity, immune  
XX stimulating or suppressing activity, haematopoiesis regulating  
XX activity, tissue growth activity, activin/inhibin activity,  
XX chemotactic/chemokinetic activity, haemostatic and thrombolytic  
XX activity, receptor/ligand activity, anti-inflammatory activity,  
XX cadherin/tumour invasion suppressor activity, tumour inhibition  
XX activity. The polynucleotide may also be useful for gene therapy.  
XX  
XX Sequence 203 BP; 59 A; 43 C; 39 G; 62 T; 0 other;  
XX

Query Match	11.1%;	Score 67.2;	DB 20;	Length 203;
Best Local Similarity	68.4%;	Pred. No. 2.3e-05;		
Matches 93;	Conservative 0;	Mismatches 43;	Indels 0;	Gaps 0;

Qy	120	GAATTTTCATCTTCAGACTGAGCTCACCAAAATGAAAGACTTATCTTGTCTTCAT	179
Db	47	GATCTCTCTTGTAGTAAAGACTGACCACTGAAAGTTTGTGTTTGTCTTAAAT	106
Qy	180	TATGGCTCATCTCTAGCATGATTAGAGCTGATTCATCTGAGAGAAAGCTCACAGAA	239
Db	107	CTTGAGCTTCATCTTTCATGACTGAGAGCTGATTCATGCAAGAGACATCATGGGTA	166
Qy	240	ACGGAAGAAACATCAT	255
Db	167	TAAAGAAATTCAT	182

RESULT 15  
AAA48965

AC AAA48965;

DT 06-OCT-2000 (first entry)

**Human basic histidine-rich protein DNA**

KW lysine-rich scatterin protein; LRSP: acidic phosphoprotein; human precipitation inhibitor; autolimmune; inflammatory disorder; AIDS; KW acthna; allergy; diabetes mellitus; fungal; bacterial infection; KW cancer; leukemia; adenocarcinoma; melanoma; ds.

**Homo sapiens**

FH	Key	Location/Qualifiers
FT	CDS	38..193

PN WO200024779-A1

PD 04-MAY-2000

PF 22-OCT-1999; 99WO-US24046.

PR 23-OCT-1998; 98US-0155209.

PA (INCY-) INCYTE PHARM INC.

PI Tang YT, Corley NC, Guegler KJ, Patterson C;

DR WPI; 2000-350699/30.  
DR P-PSDB;.AAY94528.

PT	Purified polypeptide used for treating or preventing a disorder
PT	characterized by expression or activity of lysine-rich statherin
PT	proteins -

PS Disclosure; Page 72; 75pp; English

CC The preet invention relates to human lysine-rich statherin protein (LRSP) (AA945426). The cDNA sequence encoding this protein was identified through analysis of a cDNA library of breast tumour tissue (B83T01T4). CC The LRSP sequence was found to have homology with human statherin protein (AA945427) and human basic histidine-rich protein (the present CC sequence). Human statherin is a phosphoprotein that acts as an inhibitor of precipitation of calcium phosphate salts in the oral cavity. The LRSP CC polypeptide and its antagonists may be useful for treating or preventing CC disorders associated with the activity of LRSP. Such disorders include CC autoimmune/inflammatory disorders (for example AIDS, allergies, asthma, CC diabetes mellitus), bacterial and fungal infection and cancers (such as CC leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be useful CC for diagnosis of the above disorders.

XX Sequence 491 BP; 151 A; 90 C; 87 G; 163 T; 0 other;  
SQ

Query Match	10.4%	Score 62.6;	DB 21;	Length 491;
Best Local Similarity	66.9%	Pred. No. 0.00026;		
Matches 89;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0

Oy	140	GGAGCTCACCAAAATGAAGATCTTATCTTGTCTTCATTATAGGCTCTCATCCCTACCA	195
Db	24	GGATTCAACCAACTAAGAAGTTTGTGTTTTGCTTTATCTTGCTCTCATGCTTCCA	83
Oy	200	TGATTTAGAGCTGATCTGATCTGAAGAAACGTCACAGAAACGAAAAAATCATCATGAG	255
Db	84	TGACTGGAAGCTGATTACATGCAAGAGACATCATGAGGTATATAAAGAAAAATTCATGAA	143
Oy	260	GATATTTTCAAC	272
Db	144	AGCATCATTCACA	156

Search completed: December 13, 2003, 00:19:45  
Job time : 253 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2003, 23:23:37 ; Search time 61 Seconds  
(without alignments)  
4370.417 Million cell updates/sec

Title: US-10-079-754A-4  
Perfect score: 604  
Sequence: 1 gaagattcttcagttctcata.....tatcaagcataaaaaaaaa 604

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTDS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41.2	6.8	1864	US-09-468-265-4	Sequence 4, Appl1
2	41	6.8	11469	US-09-367-895-29	Sequence 29, Appl1
3	41	6.8	11478	US-08-981-803-29	Sequence 29, Appl1
4	41	6.8	11478	US-08-983-440-29	Sequence 29, Appl1
5	40.8	6.8	6152	US-08-973-462-1	Sequence 1, Appl1
6	40.4	6.7	465	5496550-9	Patent No. 5496550
7	40.4	6.7	466	5496550-7	Patent No. 5496550
8	40.2	6.7	6152	US-08-973-462-1	Sequence 1, Appl1
9	40	6.6	789	US-09-702-705-214	Sequence 214, App
10	40	6.6	789	US-09-736-457-214	Sequence 214, App
11	39.8	6.6	4185	US-09-417-485D-7	Sequence 7, Appl1
12	39.8	6.6	6107	US-09-482-273-47	Sequence 47, Appl1
13	39.8	6.6	10640	US-09-417-485D-5	Sequence 5, Appl1
14	39.6	6.6	10640	US-09-417-485D-5	Sequence 5, Appl1
15	39	6.5	19124	US-08-487-826B-13	Sequence 13, Appl1
16	39	6.5	1664976	US-08-916-421B-1	Sequence 1, Appl1
17	38.8	6.4	19124	US-08-487-826B-13	Sequence 13, Appl1
18	38.4	6.4	989	US-08-817-926-2	Sequence 2, Appl1
19	38.4	6.4	3562	US-08-817-926-2	Sequence 2, Appl1
20	38.4	6.4	202001	US-09-734-674-3	Sequence 3, Appl1
21	38	6.3	1897	US-08-184-632-1	Sequence 1, Appl1
22	37.4	6.2	5340	US-09-627-123-21	Sequence 21, Appl1
23	37.4	6.2	5852	US-07-867-106-2	Sequence 2, Appl1
24	37	6.1	840	US-09-134-001C-1831	Sequence 1831, Ap
25	37	6.1	2621	US-08-553-619B-8	Sequence 8, Appl1
26	37	6.1	6124	US-08-213-419B-3	Sequence 3, Appl1
27	37	6.1	1664976	US-08-916-421B-1	Sequence 1, Appl1

28	36.4	6.0	1274	US-09-523-899A-1	Sequence 1, Appl1
29	36.4	6.0	152331	US-09-128-155-16	Sequence 16, Appl1
30	36.2	6.0	1173	US-09-561-322-1	Sequence 1, Appl1
31	36.2	6.0	2110	US-07-419-459-1	Sequence 1, Appl1
32	36.2	6.0	2538	US-07-839-433-1	Sequence 1, Appl1
33	36.2	6.0	8920	US-08-446-855A-1	Sequence 1, Appl1
34	36.2	6.0	8920	US-09-150-741-1	Sequence 1, Appl1
35	35.8	5.9	579	US-09-134-001C-1835	Sequence 1835, Ap
36	35.8	5.9	6265	US-09-129-112-3	Sequence 3, Appl1
37	35.8	5.9	580073	US-08-545-526D-1	Sequence 1, Appl1
38	35.6	5.9	601	US-09-820-002-13	Sequence 13, Appl1
39	35.6	5.9	2411	US-07-952-755-1	Sequence 1, Appl1
40	35.6	5.9	2411	US-08-443-679-1	Sequence 1, Appl1
41	35.6	5.9	1830121	US-09-557-884-1	Sequence 1, Appl1
42	35.6	5.9	1830121	US-09-643-890A-1	Sequence 1, Appl1
43	35.4	5.9	3001	US-09-539-333D-146	Sequence 146, App
44	35.2	5.8	633	US-09-107-532A-596	Sequence 596, App
45	35.2	5.8	1001	US-09-671-317-404	Sequence 404, App

ALIGNMENTS

RESULT 1  
US-09-468-265-4  
Sequence 4, Application US/09468265  
Patent No. 6379928  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M  
APPLICANT: Cullen, Daniel  
APPLICANT: Gray, Gregory L  
APPLICANT: Hengge, Kirk J  
APPLICANT: Lawlis, Virgil B  
TITLE OF INVENTION: Heterologous Polypeptides Expressed in Filamentous Fungi, Process  
FILE REFERENCE: A-42909-5  
CURRENT APPLICATION NUMBER: US/09/468,265  
PRIOR FILING DATE: 1999-12-10  
PRIOR APPLICATION NUMBER: 06/484,384  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 08/284,942  
PRIOR FILING DATE: 1994-08-02  
PRIOR APPLICATION NUMBER: 07/413,010  
PRIOR FILING DATE: 1989-09-25  
PRIOR APPLICATION NUMBER: 07/163,219  
PRIOR FILING DATE: 1988-02-26  
PRIOR APPLICATION NUMBER: 06/882,224  
PRIOR FILING DATE: 1986-07-07  
PRIOR APPLICATION NUMBER: 06/771,374  
PRIOR FILING DATE: 1985-08-29  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Patent version 3.1  
SEQ ID NO 4  
LENGTH: 1864  
TYPE: DNA  
ORGANISM: Emeritella nidulans  
US-09-468-265-4  
Query Match 6.8%; Score 41.2; DB 4; Length 1864;  
Best Local Similarity 46.3%; Pred. No. 0.6;  
Matches 136; Conservative 0; Mismatch 158; Indels 0; Gaps 0;  
QY 221 AAGGAAAGCTGACAGGAAACGAAACATCTATAGAGATTTTTCACAAATACGAC 280  
DB 1224 AAGGATATACGACGAGGTACATATTAATTAATAAATAAATAAAGCTACTA 1283  
QY 281 CATATCAAGCATATCACTAATATATCTCTGCTGCTTCTTAATAGCTGCTT 340  
DB 1284 AATCGAATATATTAAGAAATAGTATTTATCTAATCTTAATCTTATCTA 1343  
QY 341 AGTACTACAGACATGATTAAGAGATTTTTCACAAATGATTTTCTACTCTTCTG 400  
DB 1344 AAGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1403

```

RESULT 4
US-08-983-440-29/c
: Sequence 29, Application US/08983440
: Patent No. 6232122
GENERAL INFORMATION:
APPLICANT: POULSEN, Peter
TITLE OF INVENTION: INHIBITION OF GENE EXPRESSION
FILE REFERENCE: 674509-2003
CURRENT APPLICATION NUMBER: US/08/983,440
CURRENT FILING DATE: 1998-04-17
EARLIER APPLICATION NUMBER: 9514437.4
EARLIER FILING DATE: 1995-07-14
EARLIER APPLICATION NUMBER: PCT/EP96/03053
EARLIER FILING DATE: 1996-07-12
NUMBER OF SEQ ID NOS: 35
SOFTWARE: Patentn Ver. 2.0
SEQ ID NO 29
LENGTH: 11478
TYPE: DNA
ORGANISM: Solanum tuberosum
FEATURE:
NAME/KEY: variation
LOCATION: (1)..(11478)
OTHER INFORMATION: B stands for G or C or T/U
FEATURE:
NAME/KEY: variation
LOCATION: (1)..(11478)
OTHER INFORMATION: R stands for G or A
FEATURE:
NAME/KEY: variation
LOCATION: (1)..(11478)
OTHER INFORMATION: K stands for G or T/U
FEATURE:
NAME/KEY: variation
LOCATION: (1)..(11478)
OTHER INFORMATION: W stands for A or T/U
FEATURE:
NAME/KEY: variation
LOCATION: (1)..(11478)
OTHER INFORMATION: M stands for A or C
US-08-983-440-29

```

Query Match 6.8%; Score 41; DB 3; Length 11478;  
Best Local Similarity 51.4%; Pred. No. 1.1;  
Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 368 TTTTTCACATATTTTCTCTCTCTTTGTGTTGTAACCAATCTTCAATGAAT 427  
DB 11381 TTTGACATCTTTCTGTATATAATTTCTTTTCAATCAACCAATGAAG 11322  
QY 428 AAAACAGAAAGAAAAATCAGTCAAGTGTGCAACACATCTTGAATCAATATC 487  
DB 11321 AAAACAGAAAGAAAAATCAGTCAAGTGTGCAACACATCTTGAATGAAG 11262  
QY 488 AATATTTTAAACATATATATAGTCTCTGACATATATGTTTCTACTTTCTT 547  
DB 11261 AATCTACTGAAAGAAATATATGTAACCTTGAACCTTCAATATGACCTTCGT 11202  
QY 548 TCTCT 552  
DB 11201 CTACT 11197

RESULT 5  
US-08-973-462-1  
; Sequence 1, Application US/08973462B

; Patent No. 6191270  
; GENERAL INFORMATION:  
; APPLICANT: DRUIHE, PIERRE  
; APPLICANT: DAUBERSIES, PIERRE  
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
; FILE REFERENCE: 0660-0125-0 PCT  
; CURRENT APPLICATION NUMBER: US/08/973,462B  
; EARLIER FILING DATE: 1998-02-06  
; EARLIER APPLICATION NUMBER: PCT/FR96/00894  
; EARLIER FILING DATE: 1996-06-12  
; EARLIER APPLICATION NUMBER: FR 95/07007  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 6152  
; TYPE: DNA  
; ORGANISM: P. falciparum  
US-08-973-462-1

Query Match 6.8%; Score 40.8; DB 3; Length 6152;  
Best Local Similarity 53.0%; Pred. No. 1;  
Matches 87; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 420 AATGATTAACAAAGAAAAATCAGTCAAGTGTGCAACACATCTTGAAT 479  
DB 309 AAAAAAAAAAAAAAAAAAATCAGTATATAGTATATATATATATATATA 368  
QY 480 CAATATCAATATTTTAAACATATATATAGTCTCTGAACTATGTATGTTCTA 539  
DB 369 TAT 428  
QY 540 CTTTCTCTGTCGACTTACATGATGCTTAATAATGAT 583  
DB 429 TATATCTTTTATGCTGATATATAAGAGTTGAAAAAAT 472

RESULT 6  
5496550-9  
; Patent No. 5496550  
; APPLICANT: WALLACH, MICHAEL; PUGATSCH, THEA; MENCHER, DAVID  
; TITLE OF INVENTION: METHOD OF REDUCING THE OUTPUT OF EIMERIA  
; OCCYSTS FROM A MEMORIAL CHICK  
; NUMBER OF SEQUENCES: 10  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/108,763  
; FILING DATE: 17-AUG-1993  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 642,219  
; FILING DATE: 16-JAN-1991  
; APPLICATION NUMBER: 310,603  
; FILING DATE: 14-FEB-1989  
; APPLICATION NUMBER: 155,245  
; FILING DATE: 12-FEB-1988  
; APPLICATION NUMBER: 896,611  
; FILING DATE: 14-AUG-1986  
; SEQ ID NO: 9  
; LENGTH: 465  
5496550-9

Query Match 6.7%; Score 40.4; DB 6; Length 465;  
Best Local Similarity 51.7%; Pred. No. 0.64;  
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 425 AATAAACAAGAAAAAATCAGTCAAGTGTGCAACACATCTTGAATCAAT 484  
DB 149 AAGCAAGAAACACCTCTTATAGATGATGATTAATTAATTAATTAATTA 208  
QY 485 ATCAATATTTTAAACATATATATAGTCTCTGAACTATGTTTCTACTTTC 544  
DB 209 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 268  
QY 545 TTTTCTCTGTCATTCATGATGCTTAATTAATTAATTAATTAATTAATTA 602  
DB 269 GTGCTTTGTGATTCATGATGATGATTAATTAATTAATTAATTAATTA 326

RESULT 7  
5496550-7

; Patent No. 5496550  
; APPLICANT: WALLACH, MICHAEL; PUGATSCH, THEA; MENCHER, DAVID  
; TITLE OF INVENTION: METHOD OF REDUCING THE OUTPUT OF EIMERIA  
; OCCYSTS FROM A MEMORIAL CHICK  
; NUMBER OF SEQUENCES: 10  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/108,763  
; FILING DATE: 17-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 642,219  
; FILING DATE: 16-JAN-1991  
; APPLICATION NUMBER: 310,603  
; FILING DATE: 14-FEB-1989  
; APPLICATION NUMBER: 155,245  
; FILING DATE: 12-FEB-1988  
; APPLICATION NUMBER: 896,611  
; FILING DATE: 14-AUG-1986  
; SEQ ID NO: 7  
; LENGTH: 466  
5496550-7

Query Match 6.7%; Score 40.4; DB 6; Length 466;  
Best Local Similarity 51.7%; Pred. No. 0.64;  
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 425 AATAAACAAGAAAAAATCAGTCAAGTGTGCAACACATCTTGAATCAAT 484  
DB 150 AAGCAAGAAACACCTCTTATAGATGATGATTAATTAATTAATTAATTA 209  
QY 485 ATCAATATTTTAAACATATATATAGTCTCTGAACTATGTTTCTACTTTC 544  
DB 210 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 269  
QY 545 TTTTCTCTGTCATTCATGATGCTTAATTAATTAATTAATTAATTAATTA 602  
DB 270 GTGCTTTGTGATTCATGATGATGATTAATTAATTAATTAATTAATTA 327

RESULT 8  
US-08-973-462-1/c  
; Sequence 1, Application US/08973462B  
; Patent No. 6191270





APPLICANT: Long, David M.  
APPLICANT: Metz, Anneke M.  
APPLICANT: Love, Rueschelle A.  
TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes  
FILE REFERENCE: 47714-5009-US  
CURRENT APPLICATION NUMBER: US/09/417,485D  
CURRENT FILING DATE: 2002-06-14  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 4185  
TYPE: DNA  
ORGANISM: Plasmodium falciparum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(2382)  
OTHER INFORMATION: Partial TERT gene  
US-09-417-485D-7

Query Match  
Best Local Similarity 55.4%; Score 39.8; DB 4; Length 4185;  
Matches 77; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 387 CTACTCTTCTGTTGTTGAAACCATCTTCAATGAATAAACAAAGAAAAAAT 446  
DB 2662 CCAATCTATGTTGGACACCAAGCCAGTATATAAAAAAGAAAAATATAAAAAA 2603  
QY 447 CAGTCAAGTAGTGCACACACATCTTGGAATCAATTCATTTTAAACATATA 506  
DB 2602 TATGCAATCCAAATCATATATATTTAAAAACAATTAACAATCAAAAAA 2543  
QY 507 ATGATAGTCTCTGAACAT 525  
DB 2542 ATAAATGAATGTATATAAT 2524

RESULT 12  
US-09-482-273-47/c  
Sequence 47, Application US/09482273  
Patent No. 6534631  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 71 Human Secreted Proteins  
FILE REFERENCE: P2030P1  
CURRENT APPLICATION NUMBER: US/09/482,273  
CURRENT FILING DATE: 2000-01-13  
EARLIER APPLICATION NUMBER: PCT/US99/15849  
EARLIER FILING DATE: 1999-07-14  
EARLIER APPLICATION NUMBER: 60/092,921  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/092,922  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/092,956  
EARLIER FILING DATE: 1998-07-15  
NUMBER OF SEQ ID NOS: 267  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 47  
LENGTH: 6107  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5749)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5892)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5896)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:

NAME/KEY: SITE  
LOCATION: (5906)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5957)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5966)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-482-273-47

Query Match  
Best Local Similarity 59.1%; Score 39.8; DB 4; Length 6107;  
Matches 68; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 432 CAAAGAAAAAAATCACTCAAGTAGTGCACACACATCTTGGAATCAATATCAATA 491  
DB 2389 CAAAAAAGAAAGTAAGTAACTAGACACTGGACCTGGACATCCCAATATG 2330  
QY 492 TTTTAAACAT 546  
DB 2329 ATATTTGATTAATCTGATATTTACTGCAATATAGCATTAATCTCAAAATTT 2275

RESULT 13  
US-09-417-485D-5/c  
Sequence 5, Application US/09417485D  
Patent No. 6541202  
GENERAL INFORMATION:  
APPLICANT: Long, David M.  
APPLICANT: Metz, Anneke M.  
APPLICANT: Love, Rueschelle A.  
TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes  
FILE REFERENCE: 47714-5009-US  
CURRENT APPLICATION NUMBER: US/09/417,485D  
CURRENT FILING DATE: 2002-06-14  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 10640  
TYPE: DNA  
ORGANISM: Plasmodium falciparum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (834)..(7385)  
OTHER INFORMATION: TERT gene  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1821)..(1837)  
OTHER INFORMATION: m at position 1821 = a or c; w at position 1837 =  
OTHER INFORMATION: a or t. Xaa (amino acid) at position 330 = Leu or  
OTHER INFORMATION: Ile; Xaa at position 335 = Asp or Gly.  
US-09-417-485D-5

Query Match  
Best Local Similarity 55.4%; Score 39.8; DB 4; Length 10640;  
Matches 77; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 387 CTACTCTTCTGTTGTTGAAACCATCTTCAATGAATAAACAAAGAAAAAAT 446  
DB 7664 CCAATCTATGTTGGACACCAAGCCAGTATATAAAAAAGAAAAATATAAAAAA 7605  
QY 447 CAGTCAAGTAGTGCACACACATCTTGGAATCAATTCATTTTAAACATATA 506  
DB 7604 TATGCAATCCAAATCATATATATTTAAAAACAATTAACAATCAAAAAA 7545  
QY 507 ATGATAGTCTCTGAACAT 525  
DB 7544 ATAAATGAATGTATATAAT 7526

RESULT 14  
US-09-417-485D-5  
; Sequence 5, Application US/09417485D  
; Patent No. 6541202  
; GENERAL INFORMATION:  
; APPLICANT: Long, David M.  
; APPLICANT: Metz, Anneke M.  
; APPLICANT: Love, Ruschelle A.  
; TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes  
; FILE REFERENCE: 47714-5009-US  
; CURRENT APPLICATION NUMBER: US/09/417,485D  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 10640  
; TYPE: DNA  
; ORGANISM: Plasmodium falciparum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (834)..(7385)  
; OTHER INFORMATION: TERT gene  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1821)..(1837)  
; OTHER INFORMATION: m at position 1821 = a or c; w at position 1837 =  
; OTHER INFORMATION: a or t. Xaa (amino acid) at position 330 = Leu or  
; OTHER INFORMATION: Ile; Xaa at position 335 = Asp or Gly.  
US-09-417-485D-5

Query Match 6.6%; Score 39.6; DB 4; Length 10640;  
Best Local Similarity 57.1%; Pred. No. 2.2;  
Matches 72; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 417 TTCAATGATTAACCAAGAAAAAATCAGTCAAGTGTGCAACACACTGTTG 476  
DB 10430 TTAAGAAAAAATAAAAAAATAAACTAATTTCTATCATACGATACAAAGATAC 10489

QY 477 AACCAATATCATATTTTAAACATATATGATGCTGTAACATATGTTGTT 536  
DB 10490 TAAATATTAATAAATAATATATATATATATATATATATATATATATATATAT 10549

QY 537 CTACTT 542  
DB 10550 TTAATT 10555

RESULT 15  
US-08-487-826B-13  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,826B  
; FILING DATE: 10-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelson, Ned  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH121.001CP1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 19124 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 6.5%; Score 39; DB 2; Length 19124;  
Best Local Similarity 53.0%; Pred. No. 3.6;  
Matches 133; Conservative 0; Mismatches 110; Indels 8; Gaps 2;

QY 356 TGATTAGAGAGATTTTCAATGATTTTCTACTCTTCTGTTGAAACCATC 415  
DB 6290 TGAAT 6349

QY 416 TTCAATGATTAACCAAGAAAAAATAAAATCAGTCAAGTGTGCAACACACTGTTG 475  
DB 6350 TGAAT 6403

QY 476 GAATCAATATCATATTTTAAACATATAT--ATGATAGTCTCTGAACATATGTAATGG 533  
DB 6404 AAAAATTTTAT 6463

QY 534 TTCTACTTCTTTCTGCTGCTACATGATGATGCTTAATTAATTTGATCTACAGCA 593  
DB 6464 TTATCATTTTCTTTTGTGCTATATATATATATATATATATATATATATATATAT 6523

QY 594 TAAAAAATAA 604  
DB 6524 ACAAATAATA 6534

Search completed: December 13, 2003, 01:36:48  
Job time: 67 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2003, 00:15:48 ; Search time 278 Seconds  
(without alignments)  
7221.057 Million cell updates/sec

Title: US-10-079-754A-4

Perfect score: 604  
Sequence: 1 gaagatatttcagttctata.....tatcaagataaaaaaaaa 604

Scoring table: IDENTITY NUC  
Gapox 10.0, Gapext 1.0

Searched: 2201672 seqs, 166179599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	604	100.0	604	US-10-079-754A-4	Sequence 4, Appl1
2	465.4	77.1	505	US-10-079-754A-1	Sequence 1, Appl1
3	448.2	74.2	585	US-10-079-754A-2	Sequence 2, Appl1
4	423.4	70.1	525	US-10-079-754A-6	Sequence 6, Appl1
5	238.4	39.5	267	US-10-079-754A-15	Sequence 15, Appl1
6	238.4	39.5	267	US-10-079-623-200	Sequence 200, App
7	115	13.0	869	US-10-079-754A-3	Sequence 3, Appl1
8	87	14.4	96	US-10-079-754A-5	Sequence 5, Appl1
9	70.4	11.7	438	US-09-924-600A-5	Sequence 5, Appl1
10	70.4	11.7	438	US-09-924-340-5	Sequence 5, Appl1
11	70.4	11.7	438	US-09-927-095B-5	Sequence 5, Appl1
12	70.4	11.7	438	US-10-154-678-5	Sequence 5, Appl1
13	70.4	11.7	438	US-09-999-570-5	Sequence 5, Appl1
14	70.4	11.7	438	US-10-001-142-5	Sequence 5, Appl1
15	70.4	11.7	438	US-10-000-489-5	Sequence 5, Appl1

16	70.4	11.7	438	US-10-000-986-5	Sequence 5, Appl1
17	59.2	9.8	1521	US-10-027-632-263847	Sequence 263847, Ap
18	59.2	9.8	1521	US-10-027-632-263848	Sequence 263848, Ap
19	59.2	9.8	1521	US-10-027-632-263849	Sequence 263849, Ap
20	59.2	9.8	1521	US-10-027-632-263847	Sequence 263847, Ap
21	59.2	9.8	1521	US-10-027-632-263848	Sequence 263848, Ap
22	59.2	9.8	1521	US-10-027-632-263849	Sequence 263849, Ap
23	53.6	8.9	13449	US-10-311-455-1358	Sequence 1358, Ap
24	52	8.6	3673778	US-10-312-841-2	Sequence 2, Appl1
25	51.2	8.5	6255	US-10-311-455-934	Sequence 934, App
26	51	8.4	12507	US-10-311-455-271	Sequence 271, App
27	51	8.4	673778	US-10-312-841-1	Sequence 1, Appl1
28	49.8	8.2	7106	US-10-311-455-1451	Sequence 1451, Appl1
29	49.4	8.2	5945	US-10-311-455-57	Sequence 57, Appl1
30	49.4	8.2	16994	US-10-311-455-962	Sequence 962, App
31	49	8.1	73334	US-10-311-455-2098	Sequence 2098, App
32	47.8	7.9	6782	US-10-311-455-750	Sequence 750, App
33	47.6	7.9	6274	US-10-172-086-19	Sequence 19, Appl1
34	47.6	7.9	19659	US-10-311-455-739	Sequence 739, App
35	47.2	7.8	6155	US-10-240-453-236	Sequence 236, App
36	47	7.8	490	US-09-814-353-4552	Sequence 4552, Ap
37	47	7.8	490	US-09-814-353-10853	Sequence 10853, A
38	47	7.8	11691	US-10-311-455-2214	Sequence 2214, Ap
39	47	7.8	15732	US-10-240-453-107	Sequence 107, App
40	47	7.8	15732	US-10-239-676-95	Sequence 95, Appl1
41	46.4	7.7	6311	US-10-311-455-1936	Sequence 1936, App
42	46.2	7.6	5641	US-10-311-455-1369	Sequence 1369, Ap
43	46.2	7.6	12138	US-10-311-455-1916	Sequence 1916, Ap
44	46.2	7.6	12138	US-10-240-453-210	Sequence 210, App
45	46.2	7.6	17738	US-10-311-455-1512	Sequence 1512, Ap

## ALIGNMENTS

RESULT 1  
US-10-079-754A-4  
; Sequence 4, Application US/10079754A  
; Publication No. US20020164625A1  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Grigor, Murray R.  
; APPLICANT: Molenaar, Adrian J.  
; APPLICANT: Davis, Stephen R.  
; TITLE OF INVENTION: Compositions Isolated from Bovine  
; FILE REFERENCE: 11000.1068  
; CURRENT APPLICATION NUMBER: US/10/079, 754A  
; PRIOR FILING DATE: 2002-02-19  
; PRIOR APPLICATION NUMBER: US 09/699, 146  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US 60,162,701  
; PRIOR FILING DATE: 1999-10-29  
; PRIOR APPLICATION NUMBER: US 09/644,190  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: US 60,150,330  
; PRIOR FILING DATE: 1999-08-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 604  
; TYPE: DNA  
; ORGANISM: Bovine  
US-10-079-754A-4

Query Match 100.0%; Score 604; DB 14; Length 604;  
Best Local Similarity 100.0%; Pred. No. 4; 4e-126;  
Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAAGTATTTTACGTTCTTAATTAAGATCTCTAACTGTAATTACAAACAAATGAA 60  
DB 1 GAAGTATTTTACGTTCTTAATTAAGATCTCTAACTGTAATTACAAACAAATGAA 60

```
QY 61 GGATTTCAAGTATTTAAACACAGAGTTTCTAGCAAGAATCTCTGAGCATCAG 120
DB 61 GGATTTCAAGTATTTAAACACAGAGTTTCTAGCAAGAATCTCTGAGCATCAG 120
QY 121 AATTTCATCTTCTGACTGACCTCCACCAATATGAAAGATCTTATCTTGTCTTCAAT 180
DB 121 AATTTCATCTTCTGACTGACCTCCACCAATATGAAAGATCTTATCTTGTCTTCAAT 180
QY 181 ATGGCTCTCATCTAGGCATGATTAGAGCTGATTCATCTGAGAGAAACGTCAAGGAAA 240
DB 181 ATGGCTCTCATCTAGGCATGATTAGAGCTGATTCATCTGAGAGAAACGTCAAGGAAA 240
QY 241 CGGAAAAAATCATCATAGAGATATTTTCAACATATCCAGCCATATCAACATATCCACTA 300
DB 241 CGGAAAAAATCATCATAGAGATATTTTCAACATATCCAGCCATATCAACATATCCACTA 300
QY 301 AATTATCTCTGCGGTATCCATTTCTCTTAAAAATGCTGTTAGTAACTACAGAGATGATT 360
DB 301 AATTATCTCTGCGGTATCCATTTCTCTTAAAAATGCTGTTAGTAACTACAGAGATGATT 360
QY 361 AGAGAGATTTTTCACAAATGATTTTCTCTACTCTTCTGTTGTGAAAAACCATCTTCA 420
DB 361 AGAGAGATTTTTCACAAATGATTTTCTCTACTCTTCTGTTGTGAAAAACCATCTTCA 420
QY 421 AATGAATAAAAAAGAAAAAATCAAGTCAAGTGTGCAACAACATATCTTGAATC 480
DB 421 AATGAATAAAAAAGAAAAAATCAAGTCAAGTGTGCAACAACATATCTTGAATC 480
QY 481 AATATATCAATTTTTTAAACATAATATGATGATGCTGAACTATGTAATGGTTTCTAC 540
DB 481 AATATATCAATTTTTTAAACATAATATGATGATGCTGAACTATGTAATGGTTTCTAC 540
QY 541 TTTCTTTTCTGCTGACTTACCATGATGCTTAAATTAATGATCTATCAAGATAAAAA 600
DB 541 TTTCTTTTCTGCTGACTTACCATGATGCTTAAATTAATGATCTATCAAGATAAAAA 600
QY 601 AAAA 604
DB 601 AAAA 604
```

## RESULT 2

```
US-10-079-754A-1
; Sequence 1, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions Isolated from Bovine
; TITLE OF INVENTION: Mammary Gland and Methods for Their Use
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079, 754A
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/699,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60,162,701
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/644,190
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 60,150,330
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-754A-1
```

Query Match 77.1%; Score 465.4; DB 14; Length 505;  
Best Local Similarity 99.8%; Pred. No. 6e-95;

```
Matches 466; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 138 CTGAGCTCCACCAATATGAAAGATCTTATCTTGTCTTATTATGAGCTCTCATCTAGC 197
DB 28 CAGAGCTCCACCAATATGAAAGATCTTATCTTGTCTTATTATGAGCTCTCATCTAGC 87
QY 198 CATGATTAGAGCTGATTCATCTGAGAGAAAGCTCAAGAGAAACGAAAAAATCATCATAG 257
DB 88 CATGATTAGAGCTGATTCATCTGAGAGAAAGCTCAAGAGAAACGAAAAAATCATCATAG 147
QY 258 AGGATATTTTCAACATATCCAGCCATATCAACATATCCATATATCTCTGCGTA 317
DB 148 AGGATATTTTCAACATATCCAGCCATATCAACATATCCATATATCTCTGCGTA 207
QY 318 TCCATTTCTTAAAAATGCTGTTAGTAACTACAGAGATATGAGAGATTTTTCACAA 377
DB 208 TCCATTTCTTAAAAATGCTGTTAGTAACTACAGAGATATGAGAGATTTTTCACAA 267
QY 378 TGAATTTTCTCTACTCTTCTGTTGTGTTGAAAAACCATCTTCAATGAATAAAAAAGA 437
DB 268 TGAATTTTCTCTACTCTTCTGTTGTGTTGAAAAACCATCTTCAATGAATAAAAAAGA 327
QY 438 AAAAAAATGAGTCAAGTATGTGCAACAACATATCTTGAATGAAATATCAATATTTTAA 497
DB 328 AAAAAAATGAGTCAAGTATGTGCAACAACATATCTTGAATGAAATATCAATATTTTAA 387
QY 498 AACATATATGATGATGCTGAACTATGATGATGTTGTTCTACTTTCTTCTGTCAC 557
DB 388 AACATATATGATGATGCTGAACTATGATGATGTTGTTCTACTTTCTTCTGTCAC 447
QY 558 TTACATGATGCTTAAATTAATGATCTATCAAGATAAAAA 604
DB 448 TTACATGATGCTTAAATTAATGATCTATCAAGATAAAAA 494
```

## RESULT 3

```
US-10-079-754A-2
; Sequence 2, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions Isolated from Bovine
; TITLE OF INVENTION: Mammary Gland and Methods for Their Use
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079, 754A
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/699,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60,162,701
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/644,190
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 60,150,330
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-754A-2
```

Query Match 74.2%; Score 448.2; DB 14; Length 585;  
Best Local Similarity 99.3%; Pred. No. 4.7e-91;  
Matches 450; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 138 CTGAGCTCCACCAATATGAAAGATCTTATCTTGTCTTATTATGAGCTCTCATCTAGC 197  
DB 133 CAGAGCTCCACCAATATGAAAGATCTTATCTTGTCTTATTATGAGCTCTCATCTAGC 192

QY 198 CATGATTAGAGCTGATTCAATCTGAGAGAAAGCTCACAGAAACGGAACAAATCATCATG 257  
|  
|  
|  
DB 193 CATGATTAGAGCTGATTCAATCTGAGAGAAACGCTCACAGAAACGGAACAAATCATCATG 252  
|  
|  
|  
QY 258 AGGATATTTTCAACAAATACAGGCAATCAACGATATCCACTTAATATTCCTCTGCGGA 317  
|  
|  
|  
DB 253 AGGATATTTTCAACAAATACAGGCAATCAACGATATCCACTTAATATTCCTCTGCGGA 312  
|  
|  
|  
QY 318 TCCATTTCTTAAATAGCTGCTTAGTAATCAAGACATGATAGAGATTTTTCACAA 377  
|  
|  
|  
DB 313 TCCATTTCTTAAATAGCTGCTTAGTAATCAAGACATGATAGAGATTTTTCACAA 372  
|  
|  
|  
QY 378 TGAATTTTCTTAAATAGCTGCTTAGTAATCAAGACATGATAGAGATTTTTCACAA 437  
|  
|  
|  
DB 373 TGAATTTTCTTAAATAGCTGCTTAGTAATCAAGACATGATAGAGATTTTTCACAA 432  
|  
|  
|  
QY 438 AAAAAAATCAGTCAAGTGTGCGACAAACATCTTGAATCAATATCAATATTTTAA 497  
|  
|  
|  
DB 433 AAAAAAATCAGTCAAGTGTGCGACAAACATCTTGAATCAATATCAATATTTTAA 492  
|  
|  
|  
QY 498 AACATATATGATAGTCTCTGACTATGTAATGTTTCTACTTTCTTTCTGCTGAC 557  
|  
|  
|  
DB 493 AACATATATGATAGTCTCTGACTATGTAATGTTTCTACTTTCTTTCTGCTGAC 552  
|  
|  
|  
QY 558 TTACCATGATGCTTATATATATGATCTATCAA 590  
|  
|  
|  
DB 553 TTACCATGATGCTTATATATATGATCTATCAA 585  
|  
|  
|

RESULT 4  
US-10-079-754A-6

/ Sequence 6, Application US/10079754A  
/ Publication No. US20020164625A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Glenn, Matthew  
/ APPLICANT: Grigor, Murray R.  
/ APPLICANT: Molenaar, Adrian J.  
/ APPLICANT: Davis, Stephen R.  
/ TITLE OF INVENTION: Compositions Isolated from Bovine  
/ TITLE OF INVENTION: Mammary Gland and Methods for Their Use  
/ FILE REFERENCE: 11000.1068  
/ CURRENT APPLICATION NUMBER: US/10/079,754A  
/ PRIOR FILING DATE: 2002-02-19  
/ PRIOR APPLICATION NUMBER: US 09/699,146  
/ PRIOR FILING DATE: 2000-10-27  
/ PRIOR APPLICATION NUMBER: US 60,162,701  
/ PRIOR FILING DATE: 1999-10-29  
/ PRIOR APPLICATION NUMBER: US 09/644,190  
/ PRIOR FILING DATE: 2000-08-22  
/ PRIOR APPLICATION NUMBER: US 60,150,330  
/ PRIOR FILING DATE: 1999-08-23  
/ NUMBER OF SEQ ID NOS: 15  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 6  
/ LENGTH: 525  
/ TYPE: DNA  
/ ORGANISM: Bovine  
US-10-079-754A-6

Query Match 70.1%; Score 423.4; DB 14; Length 525;  
Best Local Similarity 93.4%; Pred. No. 1.7e-85;  
Matches 466; Conservative 0; Mismatches 1; Indels 32; Gaps 1;

QY 138 CTGAGCTCCACCAAAATATGAGATCTTTATCTTTGCTTCATATAGGCTCTCATCTAGC 197  
|  
|  
|  
DB 27 CAGGACTCCACCAAAATATGAGATCTTTATCTTTGCTTCATATAGGCTCTCATCTAGC 86  
|  
|  
|  
QY 198 CATGATTAGAGCTGATTCAATCTGAGAGAAAGCTCACAGAAACGGAACAAATCATCATG 255  
|  
|  
|  
DB 87 CATGATTAGAGCTGATTCAATCTGAGAGAAAGCTCACAGAAACGGAACAAATCATCATG 146  
|  
|  
|  
QY 256 -----AGAGATATTTTCAACAAATACGAGCCATAT 285  
|  
|  
|

DB 147 TGATAGGCTCCAGATTTCTTAATAATCAAGAGATATTTTCAACAAATACGAGCCATAT 206  
|  
|  
|  
QY 286 CAACGATATCCACTAAATATATCTCTGCGATATCCATTTCTTAAATAGCTTAGTAA 345  
|  
|  
|  
DB 207 CAACGATATCCACTAAATATATCTCTGCGATATCCATTTCTTAAATAGCTTAGTAA 266  
|  
|  
|  
QY 346 CTACAGGACATGATTAGAGAGATTTTTCACATGATATTTTCTACTCTTCTGTTGTT 405  
|  
|  
|  
DB 267 CTACAGGACATGATTAGAGAGATTTTTCACATGATATTTTCTACTCTTCTGTTGTT 326  
|  
|  
|  
QY 406 GAAAAACATCTTTCAATGATATTAACAAAGAAAAAATATGATCAAGTATGACAA 465  
|  
|  
|  
DB 327 GAAAAACATCTTTCAATGATATTAACAAAGAAAAAATATGATCAAGTATGACAA 386  
|  
|  
|  
QY 466 CACATCTTGAATCAATATATATTTTAAACATATATATGATGCTGTAAGTAT 525  
|  
|  
|  
DB 387 CACATCTTGAATCAATATATATTTTAAACATATATATGATGCTGTAAGTAT 446  
|  
|  
|  
QY 526 GTAATTTGTTCTACTTTCTTTCTCTGCTCACTTACATGATGCTTAAATATGATCT 585  
|  
|  
|  
DB 447 GTAATTTGTTCTACTTTCTTTCTCTGCTCACTTACATGATGCTTAAATATGATCT 506  
|  
|  
|  
QY 586 ATCAAGCATAAAAA 604  
|  
|  
|  
DB 507 ATCAAGCATAAAAA 525  
|  
|  
|

RESULT 5  
US-10-079-754A-15

/ Sequence 15, Application US/10079754A  
/ Publication No. US20020164625A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Glenn, Matthew  
/ APPLICANT: Grigor, Murray R.  
/ APPLICANT: Molenaar, Adrian J.  
/ APPLICANT: Davis, Stephen R.  
/ TITLE OF INVENTION: Compositions Isolated from Bovine  
/ TITLE OF INVENTION: Mammary Gland and Methods for Their Use  
/ FILE REFERENCE: 11000.1068  
/ CURRENT APPLICATION NUMBER: US/10/079,754A  
/ PRIOR FILING DATE: 2002-02-19  
/ PRIOR APPLICATION NUMBER: US 09/699,146  
/ PRIOR FILING DATE: 2000-10-27  
/ PRIOR APPLICATION NUMBER: US 60,162,701  
/ PRIOR FILING DATE: 1999-10-29  
/ PRIOR APPLICATION NUMBER: US 09/644,190  
/ PRIOR FILING DATE: 2000-08-22  
/ PRIOR APPLICATION NUMBER: US 60,150,330  
/ PRIOR FILING DATE: 1999-08-23  
/ NUMBER OF SEQ ID NOS: 15  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 15  
/ LENGTH: 267  
/ TYPE: DNA  
/ ORGANISM: Bovine  
US-10-079-754A-15

Query Match 39.5%; Score 238.4; DB 14; Length 267;  
Best Local Similarity 99.6%; Pred. No. 5e-44;  
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 138 CTGAGCTCCACCAAAATATGAGATCTTTATCTTTGCTTCATATAGGCTCTCATCTAGC 197  
|  
|  
|  
DB 28 CAGGACTCCACCAAAATATGAGATCTTTATCTTTGCTTCATATAGGCTCTCATCTAGC 87  
|  
|  
|  
QY 198 CATGATTAGAGCTGATTCAATCTGAGAGAAAGCTCACAGAAACGGAACAAATCATCATG 257  
|  
|  
|  
DB 88 CATGATTAGAGCTGATTCAATCTGAGAGAAAGCTCACAGAAACGGAACAAATCATCATG 147  
|  
|  
|  
QY 258 AGGATATTTTCAACAAATACGAGCCATATCAAGATATCACTTAATATCTCTGCGGA 317  
|  
|  
|  
DB 148 AGGATATTTTCAACAAATACGAGCCATATCAAGATATCACTTAATATCTCTGCGGA 207  
|  
|  
|

**QY**      318 TCCATTTCTTAAATGCTGCTTAGTAACCTACAGGACATGATTAGAGAGATTTTCACA 377  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
**D6**      208 TCCATTTCTTAAATGCTGCTTAGTAACCTACAGGACATGATTAGAGAGATTTTCACA 267

```

RESULT 6
US-10-079-623-200
; Sequence 200. Application US/10079623
; Publication No. US20020169302A1
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; TITLE OF INVENTION: Compositions isolated from bovine
; TITLE OF INVENTION: mammary gland and methods for their use.
; FILE REFERENCE: 11000.1044c3
; CURRENT APPLICATION NUMBER: US/10/079,623
; CURRENT FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 200
; LENGTH: 267
; TYPE: DNA
; ORGANISM: Bovine
; US-10-079-623-200

```

	Query Match	39.5%	Score 238.4	DB 14	Length 267
	Best Local Similarity	99.6%	Pred. No. 56-44		
	Matches 239	Conservative	0	Mismatches 1	Indels 0
Qy	138	CTGAGCTCCACCAATATGAGATCTTATCTTTGCTTCATTTATGGCTCATCTAGC	197		
Db	28	CAGGACTCCACCAATATGAGATCTTATCTTTGCTTCATTTATGGCTCATCTAGC	87		
Qy	198	CATGATTGACCTGATTTCATCTGGAAGAAAAGTCACAGGAAACGGAAAAATCATCARG	257		
Db	88	CATGATTGACCTGATTTCATCTGGAAGAAAAGTCACAGGAAACGGAAAAATCATCARG	147		
Qy	258	AGGATATTTTCAACACATACACAGCCATATCAACGATATCAGTAAATTTATCCCTCGCGTA	317		
Db	148	AGGATATTTTCAACACATACACAGCCATATCAACGATATCAGTAAATTTATCCCTCGCGTA	207		
Qy	318	TCCATTTTCTTAAATGCTGCTTAGTACTACAGACATGATTAGAGATTTTTCACAA	377		
Db	208	TCCATTTTCTTAAATGCTGCTTAGTACTACAGACATGATTAGAGATTTTTCACAA	267		

```

, RESULT 7
US-10-079-754A-3
, Sequence 3, Application US/10079754A
, Publication No. US20020164625A1
, GENERAL INFORMATION:
, APPLICANT: Glenn, Matthew
, APPLICANT: Grigor, Murray R.
, APPLICANT: Molenaar, Adrian J.
, TITLE OF INVENTION: Compositions Isolated from Bovine
, TITLE OF INVENTION: Mammary Gland and Methods for Their Use
, FILE REFERENCE: 11000.1068
, CURRENT APPLICATION NUMBER: US/10/079,754A
, CURRENT FILING DATE: 2002-02-19
, PRIOR APPLICATION NUMBER: US 09/699,146
, PRIOR FILING DATE: 2000-10-27
, PRIOR APPLICATION NUMBER: US 60,162,701
, PRIOR FILING DATE: 1999-10-29
, PRIOR APPLICATION NUMBER: US 09/644,190
, PRIOR FILING DATE: 2000-08-22
, PRIOR APPLICATION NUMBER: US 60,150,330
, PRIOR FILING DATE: 1999-08-23
, NUMBER OF SEQ ID NOS: 15
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 3

```

```

; LENGTH: 869
;
; TYPE: DNA
;
; ORGANISM: Bovine
US-10-079-754A-3

```

Query Match	19.0%;	Score 115;	DB 14;	Length 869;
Best Local Similarity	92.4%;	Pred. No. 4.8e-16;		
Matches 121;	Conservative 0;	Mismatches 10;	Indels 0;	Gaps 0;

  

QY	138	CTGAGCTCCACCAAAATGAAAGATCTTTATCTTTGTCCTCATTTAGSCTCTCATCTAGC	197
Db	56	CAGAGCTCCACCAAAATGAAAGATCTTTATCTTTATCTCATTTAGSCTCTCATCTAGC	115
QY	198	CATATTAGAGCTATTTCATCTGAAAGAAACGTCACAGGAAACGGAAAAACATCTATAG	257
Db	116	CATATTAGAGCTGATTTCATCTGAAAGAAACGTCACAGGAAACGGAAAAACATCTATGT	175
QY	258	AGGATATTTTC	268
Db	176	ATGATTTCCTC	186

```

RESULT 8
US-10-079-754A-5
; Sequence 5, Application US/10079754A
; Publication No. US20020164625A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; APPLICANT: Davis, Stephen R.
; TITLE OF INVENTION: Compositions Isolated from Bovine
; TITLE OF INVENTION: Mammary Gland and Methods for Their Use
; FILE REFERENCE: 11000.1068
; CURRENT APPLICATION NUMBER: US/10/079, 754A
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/6599,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60,162,701
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/644,190
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 60,150,330
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 96
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-754A-5

```

```

Query Match      14.4% Score 87; DB 14; Length 96;
Best Local Similarity 94.7%; Pred. No. 3,5e-10;
Matches 90; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      265 TTTCACAATACCGCCATATCAACGATATCCATAATTATCCTCTGGGATTCATT 324
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1   TTTCACAATACCGCCATATGAACGATATCCATAATTATCCTCTGGGATTCATTA 60

QY      325 CCTTAAATGCTGCTTGTAAGTACTACGAGACATGAT 359
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      61 TCTTAAATGCTGCTTACCACTACGAGACATGAT 95

RESULT 9
US-09-992-600A-5
; Sequence 5, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephanie
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CNM5 AND PROTEINS AND USES THEREOF

```

```
FILE REFERENCE: 91.US4.DIV
CURRENT APPLICATION NUMBER: US/09/992,600A
CURRENT FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Jpatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
FEATURE:
NAME/KEY: CDS
LOCATION: 84..317
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 318..438
FEATURE:
NAME/KEY: polyA signal
LOCATION: 397..402
FEATURE:
NAME/KEY: polyA site
LOCATION: 423..438
US-09-992-600A-5
```

```
Query Match 11.7% Score 70.4; DB 11; Length 438;
Best Local Similarity 69.9%; Pred. No. 3.8e-06;
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY 120 GAATTCATCTTTCATGACTGACCTCCACCAATATGAAGATCTTATCTTGTCTTCAT 179
DB 50 GACTCTCCTCTTGAGTAAAGAGCTCAGCAACTATGAAGTTTGTCTTGTAGT 109
QY 180 TATGCTCTCATCTCAGCATGATTTAGAGCTGATTCATCTGAAGAAAGCTCAGAGAA 239
DB 110 CTGGCTCTCATGATTTCCATGATTAAGCGCTGATTCACATGAAGAGACATCATGGGTA 169
QY 240 ACGGAAAAACATCAT 255
DB 170 TAGAAGAAAAATTCAT 185

RESULT 10
US-09-924-340-5
Sequence 5, Application US/09924340
Publication No. US20030027248A1
GENERAL INFORMATION:
APPLICANT: Bejanth, Stephane
APPLICANT: Tanaka, Hiroaki
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US2.REG
CURRENT APPLICATION NUMBER: US/09/924,340
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
```

```
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: Jpatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
NAME/KEY: CDS
LOCATION: 84..317
NAME/KEY: 3'UTR
LOCATION: 318..438
NAME/KEY: polyA signal
LOCATION: 397..402
NAME/KEY: polyA site
LOCATION: 423..438
US-09-924-340-5
```

```
Query Match 11.7% Score 70.4; DB 11; Length 438;
Best Local Similarity 69.9%; Pred. No. 3.8e-06;
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY 120 GAATTCATCTTTCATGACTGACCTCCACCAATATGAAGATCTTATCTTGTCTTCAT 179
DB 50 GACTCTCCTCTTGAGTAAAGAGCTCAGCAACTATGAAGTTTGTCTTGTAGT 109
QY 180 TATGCTCTCATCTCAGCATGATTTAGAGCTGATTCATCTGAAGAAAGCTCAGAGAA 239
DB 110 CTGGCTCTCATGATTTCCATGATTAAGCGCTGATTCACATGAAGAGACATCATGGGTA 169
QY 240 ACGGAAAAACATCAT 255
DB 170 TAGAAGAAAAATTCAT 185
```

```
RESULT 11
US-09-992-095B-5
Sequence 5, Application US/0992095B
Publication No. US20030157485A1
GENERAL INFORMATION:
APPLICANT: Bejanth, Stephane
APPLICANT: Tanaka, Hiroaki
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US4.DIV
CURRENT APPLICATION NUMBER: US/09/992,095B
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: Jpatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
NAME/KEY: CDS
LOCATION: 84..317
FEATURE:
```

NAME/KEY: 3'UTR  
LOCATION: 318..438  
FEATURE:  
NAME/KEY: polyA\_signal  
LOCATION: 397..402  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 423..438  
US-09-992-095B-5

Query Match 11.7%; Score 70.4; DB 13; Length 438;  
Best Local Similarity 69.9%; Pred. No. 3.8e-06;

Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACGACGACGACCAATATGAGATCTTATCTTTGCTTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGAGACTCAGCCACTATGAGTTTGTCTTTGCTTTAGT 109  
QY 180 TATGCTCTCATCTTCCATGATGATTAAGCTGATTCATCTGAGAGAAAGCTCAGAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTAAGCGCTGATTCATGAGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAGAAAAATTCAT 185

RESULT 12  
US-10-154-678-5

Sequence 5, Application US/10154678  
Publication No. US20030162186A1

GENERAL INFORMATION:

APPLICANT: Benjamin, Stephane

TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

FILE REFERENCE: 182. US1. REG

CURRENT APPLICATION NUMBER: US/10/154,678

CURRENT FILING DATE: 2002-10-15

PRIOR APPLICATION NUMBER: US 09/924,340

PRIOR FILING DATE: 2001-08-06

PRIOR APPLICATION NUMBER: US 60/305,456

PRIOR FILING DATE: 2001-07-13

PRIOR APPLICATION NUMBER: US 60/302,277

PRIOR FILING DATE: 2001-06-29

PRIOR APPLICATION NUMBER: US 60/298,698

PRIOR FILING DATE: 2001-06-15

PRIOR APPLICATION NUMBER: US 60/293,574

PRIOR FILING DATE: 2001-05-25

NUMBER OF SEQ ID NOS: 112

SOFTWARE: JPatent

SEQ ID NO 5

LENGTH: 438

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: 5'UTR

LOCATION: 1..83

FEATURE:

NAME/KEY: CDS

LOCATION: 84..317

FEATURE:

NAME/KEY: 3'UTR

LOCATION: 318..438

FEATURE:

NAME/KEY: polyA\_signal

LOCATION: 397..402

FEATURE:

NAME/KEY: polyA\_site

LOCATION: 423..438

US-10-154-678-5

Query Match 11.7%; Score 70.4; DB 13; Length 438;  
Best Local Similarity 69.9%; Pred. No. 3.8e-06;

Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACGACGACGACCAATATGAGATCTTATCTTTGCTTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGAGACTCAGCCACTATGAGTTTGTCTTTGCTTTAGT 109  
QY 180 TATGCTCTCATCTTCCATGATGATTAAGCTGATTCATCTGAGAGAAAGCTCAGAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTAAGCGCTGATTCATGAGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAGAAAAATTCAT 185

Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
QY 120 GAATTCATCTTTCATGACGACGACGACCAATATGAGATCTTATCTTTGCTTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGAGACTCAGCCACTATGAGTTTGTCTTTGCTTTAGT 109  
QY 180 TATGCTCTCATCTTCCATGATGATTAAGCTGATTCATCTGAGAGAAAGCTCAGAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTAAGCGCTGATTCATGAGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAGAAAAATTCAT 185

Query Match 11.7%; Score 70.4; DB 13; Length 438;  
Best Local Similarity 69.9%; Pred. No. 3.8e-06;

Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACGACGACGACCAATATGAGATCTTATCTTTGCTTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGAGACTCAGCCACTATGAGTTTGTCTTTGCTTTAGT 109  
QY 180 TATGCTCTCATCTTCCATGATGATTAAGCTGATTCATCTGAGAGAAAGCTCAGAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTAAGCGCTGATTCATGAGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAGAAAAATTCAT 185

RESULT 13  
US-09-999-570-5

Sequence 5, Application US/09999570  
Publication No. US20030170628A1

GENERAL INFORMATION:

APPLICANT: Benjamin, Stephane

TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

FILE REFERENCE: G-091US08DIY

CURRENT APPLICATION NUMBER: US/09/999,570

CURRENT FILING DATE: 2001-06-14

PRIOR APPLICATION NUMBER: US 09/924,340

PRIOR FILING DATE: 2001-08-06

PRIOR APPLICATION NUMBER: PCT/IB01/01715

PRIOR FILING DATE: 2001-08-06

PRIOR APPLICATION NUMBER: US 60/305,456

PRIOR FILING DATE: 2001-07-13

PRIOR APPLICATION NUMBER: US 60/302,277

PRIOR FILING DATE: 2001-06-29

PRIOR APPLICATION NUMBER: US 60/298,698

PRIOR FILING DATE: 2001-06-15

PRIOR APPLICATION NUMBER: US 60/293,574

PRIOR FILING DATE: 2001-05-25

NUMBER OF SEQ ID NOS: 112

SOFTWARE: JPatent

SEQ ID NO 5

LENGTH: 438

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: 5'UTR

LOCATION: 1..83

FEATURE:

NAME/KEY: CDS

LOCATION: 84..317

FEATURE:

NAME/KEY: 3'UTR

LOCATION: 318..438

FEATURE:

NAME/KEY: polyA\_signal

LOCATION: 397..402

FEATURE:

NAME/KEY: polyA\_site

LOCATION: 423..438

US-09-999-570-5

Query Match 11.7%; Score 70.4; DB 13; Length 438;  
Best Local Similarity 69.9%; Pred. No. 3.8e-06;

Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTCATGACGACGACGACCAATATGAGATCTTATCTTTGCTTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGAGACTCAGCCACTATGAGTTTGTCTTTGCTTTAGT 109  
QY 180 TATGCTCTCATCTTCCATGATGATTAAGCTGATTCATCTGAGAGAAAGCTCAGAGAA 239  
DB 110 CTGGCTCTCATGATTTCCATGATTAAGCGCTGATTCATGAGAAAGACATCATGGGTA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAGAAAAATTCAT 185



QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAAGAAAATTCAT 185

## RESULT 14

US-10-001-142-5  
; Sequence 5, Application US/10001142  
; Publication No. US20030198954A1  
; GENERAL INFORMATION:  
; APPLICANT: Benjamin, Stephanie  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 91.US7.DIV  
; CURRENT APPLICATION NUMBER: US/10/001,142  
; CURRENT FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: JPatent  
; SEQ ID NO 5  
; LENGTH: 438  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 1..83  
; NAME/KEY: CDS  
; LOCATION: 84..317  
; NAME/KEY: 3'UTR  
; LOCATION: 318..438  
; NAME/KEY: polyA\_signal  
; LOCATION: 397..402  
; NAME/KEY: polyA\_site  
; LOCATION: 423..438  
US-10-001-142-5

Query Match 11.7%; Score 70.4; DB 13; Length 438;  
Best Local Similarity 69.9%; Pred. No. 3.8e-06;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTGATGATGAGCTCCACCAATATGAAGATCTTATTTGCTTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGAGCTCAGCAACTATGAAGTTTGTCTTGCTTTAGT 109  
QY 180 TATGCTCTCATCTAGAGCTGATTAAGAGCTGATTCATCTGAAGAAAGCTCAGAGAA 239  
DB 110 CTGGCTCTCATGATTCATGATTAAGAGCTGATTCATCTGAAGAAAGCTCAGAGAA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAAGAAAATTCAT 185

RESULT 15  
US-10-000-489-5  
; Sequence 5, Application US/10000489  
; Publication No. US20030092011A1  
; GENERAL INFORMATION:  
; APPLICANT: Benjamin, Stephanie  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 91.US6.DIV  
; CURRENT APPLICATION NUMBER: US/10/000,489  
; CURRENT FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: JPatent  
; SEQ ID NO 5  
; LENGTH: 438  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 1..83  
; NAME/KEY: CDS  
; LOCATION: 84..317  
; NAME/KEY: 3'UTR  
; LOCATION: 318..438  
; NAME/KEY: polyA\_signal  
; LOCATION: 397..402  
; NAME/KEY: polyA\_site  
; LOCATION: 423..438  
US-10-000-489-5

Query Match 11.7%; Score 70.4; DB 15; Length 438;  
Best Local Similarity 69.9%; Pred. No. 3.8e-06;  
Matches 95; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 120 GAATTCATCTTTGATGATGAGCTCCACCAATATGAAGATCTTATTTGCTTCAT 179  
DB 50 GACTCTCTCTTGAGTAAAGAGCTCAGCAACTATGAAGTTTGTCTTGCTTTAGT 109  
QY 180 TATGCTCTCATCTAGAGCTGATTAAGAGCTGATTCATCTGAAGAAAGCTCAGAGAA 239  
DB 110 CTGGCTCTCATGATTCATGATTAAGAGCTGATTCATCTGAAGAAAGCTCAGAGAA 169  
QY 240 ACGGAAAAACATCAT 255  
DB 170 TAGAAGAAAATTCAT 185

Search completed: December 13, 2003, 01:41:34  
Job time : 284 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2003, 23:18:52 ; Search time 1865 Seconds  
(without alignments)  
7871.262 Million cell updates/sec

Title: US-10-079-754A-4

Perfect score: 604  
Sequence: 1 gaagatatttcagttctata.....tatcaagcataaaaaaaaa 604

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estinu:\*  
4: em\_estmu:\*  
5: em\_estcov:\*  
6: em\_estcpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gp\_est1:\*  
10: gp\_est2:\*  
11: gp\_hic:\*  
12: gp\_est3:\*  
13: gp\_est4:\*  
14: gp\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gp\_gss1:\*  
29: gp\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82.2	13.6	553	14	CB985810 AGENCOURT
2	78.6	13.0	546	14	CB957544 AGENCOURT
3	78.6	13.0	555	14	CB986888 AGENCOURT
4	78.6	13.0	792	14	CB958289 AGENCOURT

5	78.6	13.0	793	14	CB984475 AGENCOURT
6	78.6	13.0	831	14	CB985521 AGENCOURT
7	78.6	13.0	929	14	CB987745 AGENCOURT
8	77.6	12.8	583	2	BSM076012
9	77.6	12.8	622	2	BSM075901
10	77.6	12.8	773	14	CB987175 AGENCOURT
11	77	12.7	312	9	AA376700 EST89157
12	77	12.7	327	9	AA382524 EST95743
13	77	12.7	330	9	AA376704 EST89161
14	77	12.7	348	9	AA376689 EST99145
15	77	12.7	390	13	BX283872 BX283872
16	77	12.7	527	14	CB986478 AGENCOURT
17	77	12.7	533	14	CB985159 AGENCOURT
18	77	12.7	540	14	CB957853 AGENCOURT
19	77	12.7	543	14	CB958523 AGENCOURT
20	77	12.7	543	14	CB987834 AGENCOURT
21	77	12.7	544	14	CB957223 AGENCOURT
22	77	12.7	544	14	CB959166 AGENCOURT
23	77	12.7	545	14	CB958545 AGENCOURT
24	77	12.7	545	14	CB959029 AGENCOURT
25	77	12.7	546	14	CB956307 AGENCOURT
26	77	12.7	546	14	CB956449 AGENCOURT
27	77	12.7	546	14	CB957121 AGENCOURT
28	77	12.7	546	14	CB957972 AGENCOURT
29	77	12.7	546	14	CB985184 AGENCOURT
30	77	12.7	546	14	CB986281 AGENCOURT
31	77	12.7	546	14	CB986417 AGENCOURT
32	77	12.7	546	14	CB987123 AGENCOURT
33	77	12.7	546	14	CB987595 AGENCOURT
34	77	12.7	547	14	CB955720 AGENCOURT
35	77	12.7	547	14	CB956188 AGENCOURT
36	77	12.7	547	14	CB956487 AGENCOURT
37	77	12.7	547	14	CB956536 AGENCOURT
38	77	12.7	547	14	CB956860 AGENCOURT
39	77	12.7	547	14	CB956877 AGENCOURT
40	77	12.7	547	14	CB957224 AGENCOURT
41	77	12.7	547	14	CB957280 AGENCOURT
42	77	12.7	547	14	CB957414 AGENCOURT
43	77	12.7	547	14	CB957440 AGENCOURT
44	77	12.7	547	14	CB957523 AGENCOURT
45	77	12.7	547	14	CB957940 AGENCOURT

#### ALIGNMENTS

RESULT 1  
CB985810 553 bp mRNA linear EST 01-MAY-2003  
LOCUS AGENCOURT.13647469 NIH MGC.184 Homo sapiens cDNA clone  
DEFINITION IMAGE:30329983 5', mRNA sequence.  
ACCESSION CB985810  
VERSION CB985810.1 GI:30280334  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Buxaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
1 (bases 1 to 553)  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palokovits  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDCM143 row: b column: 08  
High quality sequence stop: 553.

**FEATURES**  
**source**

```

FEATURES
SOURCE
1..553
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30329983"
/lab_host="DH10B (TI phase-resistant)"
/clone_1fb="NIH MGC_184"
/note="Organ: Pooled; Glandular; Vector: pDNR-Lib; Site: 1:
5' (ggccattatggcc); Site 2: 5' (ggccgctggcc);
Library is oligo-dr primed and directionally cloned. cDNA
was prepared from a glandular pool of tissues from thyroid
parathyroid, adrenal, cortex and pineal gland. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCCGAGCGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.38
Kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
171 a 117 c 81 g 185 t 1 others

```

**source**

```

SOURCE
1. 546
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30352504"
/lab_host="DH10B (T1 phage-resistant)"
/clone_id="NIH MGC 184"
/notes="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site_1:
SfiI (ggcccatatggcc); Site_2: SfiI (ggccgcctcgcc);
Library is oligo-dT primed and directionally cloned. cDNA
was prepared from a glandular pool of tissues from thyroid,
parathyroid, adrenal, cortex and pineal gland. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATATGCG-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCCGAGCGCGCGAGATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.38
kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
166 a 103 c 87 g 185 t 1 others

```



CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
 Plate: NDCM135 row: h column: 14  
 High quality sequence stop: 486.  
 Location/Qualifiers

## FEATURES

source

```

1. 793
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30327061"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_184"
/notes="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site_1:
SfiI (ggccatagggc); Site_2: SfiI (ggccgctgggc);
Library is oligo-dt primed and directionally cloned. CDNA
was prepared from a glandular pool of tissues from thyroid,
parathyroid, adrenal, cortex and pineal gland. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCCATTTATGCGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGCGCGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.38
kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."

```

BASE COUNT 230 a 168 c 152 g 242 t 1 others  
 ORIGIN

Query Match 13.0%; Score 78.6; DB 14; Length 793;  
 Best Local Similarity 60.2%; Pred. No. 0.074;

Matches 151; Conservative 0; Mismatches 94; Indels 6; Gaps 1;

```

QY 120 GAATTCATCTTTCATGACGCTGACGACCAATATGAAATCTTTATCTTGTCTTCAT 179
    |||
DB 41 GACTCTCCTCTTGATGAAAGGACTCAGCCAACTATGAAGTTTGTCTTGTCTTAACT 100
    |||
QY 180 TATGGCTCTCATCTTACGATGATTAGAGCTGATTCATCTGAAGAGAAAGCTCAGAGAA 239
    |||
DB 101 CTGGCTCTCATGATTTTCATGATTAAGGCTGATTCATGAAAGAGACATCTATGGGTA 160
    |||
QY 240 ACGGAAAAACATCATGAGAGATTTTCAACAATACAGCATATCAAGATATCCACT 299
    |||
DB 161 TAGAGAAAATTCAT-----GAAAAGCATCATTCACATCCAGAAATTTTCATTTTATGG 214
    |||
QY 300 AATTATCTCTCCGCGATCCATTTCTTAAATGCTGCTTAGTACTACAGACATGAT 359
    |||
DB 215 GGACTATGATCAAAATTTATCTATATGACAATGATATCTTAGTATCATGGGCGATGAT 274
    |||
QY 360 TAGAGAGATTT 370
    |||
DB 275 TATAGAGGTTT 285
    |||

```

RESULT 6  
 CB985521  
 LOCUS 831 bp mRNA linear EST 01-MAY-2003  
 DEFINITION AGENCOURT 13652179 NIH MGC 184 Homo sapiens cDNA clone  
 IMAGE:30327553 5', mRNA sequence.  
 CB985521  
 ACCESSION CB985521.1 GI:30280045  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 831)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished

## COMMENT

Contact: Robert Strauberg, Ph.D.  
 Email: [cgaps-remail.llnl.gov](mailto:cgaps-remail.llnl.gov)  
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
 Plate: NDCM136 row: m column: 02  
 High quality sequence stop: 177.  
 Location/Qualifiers

## FEATURES

source

```

1. 831
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30327553"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_184"
/notes="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site_1:
SfiI (ggccatagggc); Site_2: SfiI (ggccgctgggc);
Library is oligo-dt primed and directionally cloned. CDNA
was prepared from a glandular pool of tissues from thyroid,
parathyroid, adrenal, cortex and pineal gland. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCCATTTATGCGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGCGCGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.38
kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."

```

BASE COUNT 205 a 241 c 112 g 273 t  
 ORIGIN

Query Match 13.0%; Score 78.6; DB 14; Length 831;  
 Best Local Similarity 60.2%; Pred. No. 0.072;

Matches 151; Conservative 0; Mismatches 94; Indels 6; Gaps 1;

```

QY 120 GAATTCATCTTTCATGACGCTGACGACCAATATGAAATCTTTATCTTGTCTTCAT 179
    |||
DB 41 GACTCTCCTCTTGATGAAAGGACTCAGCCAACTATGAAGTTTGTCTTGTCTTAACT 100
    |||
QY 180 TATGGCTCTCATCTTACGATGATTAGAGCTGATTCATCTGAAGAGAAAGCTCAGAGAA 239
    |||
DB 101 CTGGCTCTCATGATTTTCATGATTAAGGCTGATTCATGAAAGAGACATATGGGTA 160
    |||
QY 240 ACGGAAAAACATCATGAGAGATTTTCAACAATACAGCATATCAAGATATCCACT 299
    |||
DB 161 TAGAGAAAATTCAT-----GAAAAGCATCATTCACATCCAGAAATTTTCATTTTATGG 214
    |||
QY 300 AATTATCTCTCCGCGATCCATTTCTTAAATGCTGCTTAGTACTACAGACATGAT 359
    |||
DB 215 GGACTATGATCAAAATTTATCTATATGACAATGATATCTTAGTATCATGGGCGATGAT 274
    |||
QY 360 TAGAGAGATTT 370
    |||
DB 275 TATAGAGGTTT 285
    |||

```

RESULT 7  
 CB987745  
 LOCUS 929 bp mRNA linear EST 01-MAY-2003  
 DEFINITION AGENCOURT 13631497 NIH\_MGC\_184 Homo sapiens cDNA clone  
 IMAGE:30327757 5', mRNA sequence.  
 CB987745  
 ACCESSION CB987745.1 GI:30282265  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 929)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgc@bbs-remail.nih.gov

Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
 cDNA Library Preparation: CLOUTCH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNL at:  
<http://image.llnl.gov>  
 Plate: NDCM137 row: e column: 14  
 High quality sequence stop: 386.

FEATURES  
 source 1..929  
 location/Qualifiers

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30327757"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIH MGC 184"  
 /note="Torgan: Pooled-glandular; Vector: pDNR-LIB; Site\_1:  
 Still (ggccatcggcc); Site\_2: Still (ggccgctcggcc);  
 library is Oligo-dT primed and directionally cloned. cDNA  
 was prepared from a glandular pool of tissues from thyroid,  
 parathyroid, adrenal, cortex and pineal gland. 5' and 3'  
 adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CACGGCCATTGAGCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCCGAGGCGGCCGACATG-(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.38  
 kb (range 0.60-3.5 kb). 15/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH-MGC Library."  
 BASE COUNT 269 a 199 c 185 g 274 t 2 others  
 ORIGIN

Query Match 13.0%; Score 78.6; DB 14; Length 929;  
 Best Local Similarity 60.2%; Pred. No. 0.069; Indels 6; Gaps 1;  
 Matches 151; Conservative 0; Mismatches 94;

QY 120 GAATTCATCTTTCATGACTGACCTCCACCAATATGAAGATCTTATCTTCTTCAT 179  
 DB 21 GACTCTCCCTTGAGTAAGAGACTGACCACTATGAGATTTTGTCTTTGCTTACT 80  
 QY 180 TATGGCTCATCTTACCATGATTTAGAGCTGATTCATCTGAAGAAAGTCAAGAA 239  
 DB 81 CTGGCTCTCATATTTCCATGATTTAGCGCTGATTCACATGAAGAGACATCATGGGTA 140  
 QY 240 ACGGAAAAAATCATATAGAGATTTTGAACAATACGACCATATCAACGATATCCACT 299  
 DB 141 TGAAGAAAAATTCAT-----GAAAGCATATTCATCAGAAATTTCCATTTTATGG 194  
 QY 300 AATATTCCTCTGCTATCATCTTCTTAAATGCTGTAGTACTACAGACATGAT 359  
 DB 195 GGACTATGATCAATATATCTATATGACATGATATCTTATGATATCATGGGATGAT 254  
 QY 360 TAGAGAGATTT 370  
 DB 255 TATAGAGGTTT 265

RESULT 8

HSN076012 standard; RNA; EST; 583 BP.

XX AC BX485825;  
 XX SV BX485825.1  
 XX DT 09-MAY-2003 (rel. 75, Created)  
 DT 09-MAY-2003 (rel. 75, Last updated, Version 1)

XX DE Homo sapiens mRNA; EST DKFZp686L09248\_r1 (from clone DKFZp686L09248)  
 XX EST: expressed sequence tag.  
 XX OS Homo sapiens (human)  
 XX OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;  
 XX Eutheria; Primates; Catarrhini; Homiidae; Homo.
 XX RN [1]  
 RP 1-583  
 RA Oltewaelder B., Obermaier B., Deutschenbaur S., Mewes H.W., Weil B.,  
 RA Amdt C., Osanger A., Fobo G., Han M., Wiemann S.;  
 RT Submitted (07-MAY-2003) to the EMBL/Genbank/DBJ databases.  
 RL MIBS, Ingolstaedter Landstr.1, D-85764 Neunherberg, GERMANY

CC This is the 5' sequence of the clone insert  
 CC Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 CC Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
 CC sequenced by Medigenomix (Martinsried/Germany) within the cDNA  
 CC sequencing consortium of the German Genome Project.  
 CC No si sequence available.  
 CC This clone (DKFZp686L09248) is available at the RZPD in Berlin.  
 CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,  
 CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

Key location/Qualifiers

FT source 1..583  
 FT /db\_xref="taxon:9606"  
 FT /mol\_type="mRNA"  
 FT /organism="Homo sapiens"  
 FT /clone="DKFZp686L09248"  
 FT /clone\_lib="686 (synonym: h1cc3). Vector pSport1\_Sfi; host  
 FT DH10B; sites Sfi1 + Sfi1B"  
 FT /dev\_stage="adult"  
 FT /tissue\_type="cDNA-collection"

Sequence 583 BP; 153 A; 98 C; 119 G; 212 T; 1 other;

Query Match 12.8%; Score 77.6; DB 2; Length 583;  
 Best Local Similarity 56.2%; Pred. No. 0.12;  
 Matches 189; Conservative 0; Mismatches 139; Indels 8; Gaps 2;

QY 120 GAATTCATCTTTCATGACTGACCTCCACCAATATGAAGATCTTATCTTCTTCAT 179  
 DB 38 GACTCTCCCTTGAGTAAGAGACTGACCACTATGAGATTTTGTCTTTGCTTACT 97  
 QY 180 TATGGCTCATCTTACCATGATTTAGAGCTGATTCATCTGAAGAAAGTCAAGAA 239  
 DB 98 CTGGCTCTCATATTTCCATGATTTAGCGCTGATTCATGAAGAGACATCATGGGTA 157  
 QY 240 ACGGAAAAAATCATATAGAGATTTTGAACAATACGACCATATCAACGATATCCACT 299  
 DB 158 TGAAGAAAAATTCAT-----GAAAGCATATTCATCAGAAATTTCCATTTTATGG 211  
 QY 300 AATATTCCTCTGCTATCATCTTCTTAAATGCTGTAGTACTACAGACATGAT 359  
 DB 212 GGACTATGATCAATATATCTATATGACAAATGATATCTTATGATATCATGGGATGAT 271  
 QY 360 TAGAGAGATTTTTCACATGATTTTCTACCTCTTCTGTTGTGTGAAGCAATCTTTC 419  
 DB 272 TATAGAG--GTAGTGAAGCTTATGTTGTTCTTCTTGAAGAGTCAAGTCAAGT 329  
 QY 420 AATGATATTAACAAAGAAAAAATATCAGTCAAGT 455  
 DB 330 TAAAAAAGCCATAGCTAAGCAACCATTCAGT 365

RESULT 9

HSN075901 standard; RNA; EST; 622 BP.







/dev\_stages="adult"  
/clone\_1ib="Testis 1"  
/note="Organ: testis; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"

BASE COUNT 95 a 67 c 53 g 112 t

ORIGIN

Query Match 12.7%; Score 77; DB 9; Length 327;  
Best Local Similarity 59.8%; Pred. No. 0.18;  
Matches 150; Conservative 0; Mismatches 95; Indels 6; Gaps 1;

QY 120 GAATTCATCTTCATGATGAGTGGAGTCCCAAAATATGAGATCTTTATCTTTCTTCAT 179  
DB 24 GACTCTCTCTGATGATGAGTGGAGTCCCAAAATATGAGATCTTTATCTTTCTTCAT 83  
QY 180 TATGGCTCTCATCTGATGAGTGGAGTCCCAAAATATGAGATCTTTATCTTTCTTCAT 239  
DB 84 CTGGCTCTCATCTGATGAGTGGAGTCCCAAAATATGAGATCTTTATCTTTCTTCAT 143  
QY 240 ACCGAAAAAATCATATGAGATGATTTTCAACAATATGAGATCTTTATCTTTCTTCAT 299  
DB 144 TAGAGAAAAATTCAT-----GAAAGCATCATTCATCATGAGATTTTCATTTATGG 197  
QY 300 AAATATCTCTGCGATTCATTTCTTAAATGCTGCTTATGATGATGAGATCTTTATCTTTCTTCAT 359  
DB 198 GCACTATGATCAAAATATCTATATGATGATGATGATGATGATGATGATGATGATGATGAT 257  
QY 360 TAGAGAGATTT 370  
DB 258 TATAGAGGTTT 268

RESULT 13  
AA376704  
LOCUS AA376704 330 bp mRNA linear EST 21-APR-1997  
DEFINITION EST89161 Salivary gland Homo sapiens cDNA 5' end similar to  
hstactin 1, mRNA sequence.  
ACCESSION AA376704  
VERSION AA376704.1 GI:2029022  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult,  
'C.J., Lee, N.H., Kirkness, E.F., Weissbrock, K.G., Gockyne, J.D., White,  
'O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Claydon, R.A.,  
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald,  
L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodok, A.,  
Graham, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,  
Kelley, J.C., Liu, L.-I., Marmore, S.M., Merrick, J.M.,  
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligriano, S.M.,  
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,  
Small, K.V., Spriggs, T.A., Utecherback, T.R., Weidman, J.F., Li, Y.,  
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
Dimke, D., Feng, D.-F., Ferris, A., Fischer, C., Hastings, G.A., He, W.W.,  
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,  
Kunsch, C., Hungjun, J., Li, H., Weisner, P.S., Olsen, H., Raymond, L.,  
Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon,  
'M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and  
Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)  
MEDLINE 96026280  
PUBMED 7566098

COMMENT  
Other ESTs: EST89160 THC102726  
Contact: Kerlavage, AR  
Bioinformatics for Genomic Research  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056

Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/cdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
source  
1.330  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="ATCC (inhost):181119"  
/db\_xref="taxon:9606"  
/dev\_stages="adult"  
/clone\_1ib="Salivary gland"  
/note="Organ: salivary gland; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 94 a 66 c 54 g 115 t 1 others

ORIGIN

Query Match 12.7%; Score 77; DB 9; Length 330;  
Best Local Similarity 59.8%; Pred. No. 0.18;  
Matches 150; Conservative 0; Mismatches 95; Indels 6; Gaps 1;

QY 120 GAATTCATCTTCATGATGAGTGGAGTCCCAAAATATGAGATCTTTATCTTTCTTCAT 179  
DB 16 GACTCTCTCTGATGATGAGTGGAGTCCCAAAATATGAGATCTTTATCTTTCTTCAT 75  
QY 180 TATGGCTCTCATCTGATGAGTGGAGTCCCAAAATATGAGATCTTTATCTTTCTTCAT 239  
DB 76 CTGGCTCTCATCTGATGAGTGGAGTCCCAAAATATGAGATCTTTATCTTTCTTCAT 135  
QY 240 ACCGAAAAAATCATATGAGATGATTTTCAACAATATGAGATCTTTATCTTTCTTCAT 299  
DB 136 TAGAGAAAAATTCAT-----GAAAGCATCATTCATCATGAGATTTTCATTTATGG 189  
QY 300 AAATATCTCTGCGATTCATTTCTTAAATGCTGCTTATGATGATGAGATCTTTATCTTTCTTCAT 359  
DB 190 GCACTATGATCAAAATATCTATATGATGATGATGATGATGATGATGATGATGATGATGAT 249  
QY 360 TAGAGAGATTT 370  
DB 250 TATAGAGGTTT 260

RESULT 14  
AA376689  
LOCUS AA376689 348 bp mRNA linear EST 21-APR-1997  
DEFINITION EST89145 Salivary gland Homo sapiens cDNA 5' end similar to  
hstactin 1, mRNA sequence.  
ACCESSION AA376689  
VERSION AA376689.1 GI:2029007  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult,  
'C.J., Lee, N.H., Kirkness, E.F., Weissbrock, K.G., Gockyne, J.D., White,  
'O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Claydon, R.A.,  
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald,  
L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodok, A.,  
Graham, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,  
Kelley, J.C., Liu, L.-I., Marmore, S.M., Merrick, J.M.,  
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligriano, S.M.,  
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,  
Small, K.V., Spriggs, T.A., Utecherback, T.R., Weidman, J.F., Li, Y.,  
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
Dimke, D., Feng, D.-F., Ferris, A., Fischer, C., Hastings, G.A., He, W.W.,  
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,  
Kunsch, C., Hungjun, J., Li, H., Weisner, P.S., Olsen, H., Raymond, L.,  
Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon,  
'M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and  
Venter, J.C.



**THIS PAGE BLANK (USPTO)**